DATA PROCESSING METHOD AND APPARATUS, AND DEVICE, COMPUTER-READABLE STORAGE MEDIUM AND COMPUTER PROGRAM PRODUCT

RELATED APPLICATION

[0001] This application is based on and claims priority to Chinese Patent Application No. 202410897362.9, filed on July 5, 2024, which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

[0002] This application relates to the field of internet technologies, and in particular, to a data processing method and apparatus, a device, a computer-readable storage medium, and a computer program product.

BACKGROUND

[0003] With continuous development of internet technologies, various applications have emerged rapidly, and users can access a wide variety of content, such as browsing news, watching videos, and reading novels, through these applications.

[0004] To achieve high-quality experience of an information stream (a feed stream, also referred to as a content stream), in the related art, a terminal device initiates a preloading request before a user consumes content, so that an application backend performs online content recommendation based on the preloading request, and the terminal device can perform online content preloading, that is, preload online content. However, performing online content preloading needs to perform a complete recommendation procedure, resulting in high consumption of computing resources. In addition, because the preloading request is initiated before the user browses the content, and not every piece of online recommended content corresponding to the preloading request is used by the user, a large quantity of ineffective preloading requests exist. Apparently, according to the related art, utilization of online recommended content is low, and timeliness of the recommended content is poor. Consequently, freshness and attractiveness of content recommended to users are reduced, resulting in poor user stickiness, and huge waste of computing resources and storage resources of an application backend.

SUMMARY

[0005] Embodiments of this application provide a data processing method and apparatus, a device, a computer-readable storage medium, and a computer program product, to reduce occupation of computing resources and improve timeliness of content recommendation.

[0006] An embodiment of this application provides a data processing method, applied to a computer device, including:

[0007] obtaining a preloading request sent by the terminal device, and obtaining, based on the preloading request, a first preloading usage probability corresponding to the terminal device, the first preloading usage probability indicating a probability that a target object corresponding to the terminal device interacts with content provided by a preloading function, and the preloading function being configured for responding to the preloading request; and

[0008] generating an online recommended content set for the terminal device if the first preloading usage probability is greater than or equal to an online recommendation threshold, and generating a target preloaded content list based on the online recommended content set, a generation timestamp of the online recommended content set being later than a generation timestamp of the preloading request.

[0009] An embodiment of this application provides a data processing apparatus, including:

[0010] an obtaining module, configured to: obtain a preloading request sent by a terminal device; and obtain, based on the preloading request, a first preloading usage probability corresponding to the terminal device, the first preloading usage probability indicating a probability that a target object corresponding to the terminal device interacts with content provided by a preloading function, and the preloading function being configured for responding to the preloading request; and

[0011] a generation module, further configured to: generate an online recommended content set for the terminal device if the first preloading usage probability is greater than or equal to an online recommendation threshold; and generate a target preloaded content list based on the online recommended content set, a generation timestamp of the online recommended content set being later than a generation timestamp of the preloading request.

[0012] An embodiment of this application provides a computer device, including: a processor, a memory, and a network interface,

[0013] the processor being connected to the memory and the network interface, the network interface being configured to provide a data communication function, the memory being configured

to store a computer program, and the processor being configured to invoke the computer program to cause the computer device to perform the method in embodiments of this application.

[0014] According to an aspect, an embodiment of this application provides a computer-readable storage medium, having a computer program stored thereon, the computer program being loaded by a processor to perform the method in embodiments of this application.

[0015] An embodiment of this application provides a computer program product, including a computer program, the computer program being stored in a computer-readable storage medium, a processor of a computer device reading the computer program from the computer-readable storage medium, and the processor executing the computer program to cause the computer device to perform the method in embodiments of this application.

[0016] It can be learned from the above that this embodiment of this application provides a resource allocation method based on the first preloading usage probability. For a terminal device corresponding to the first preloading usage probability that is greater than or equal to the online recommendation threshold, an online recommended content set generated in real time is provided, and a preloaded content list generated in real time can further be provided. In other words, the first preloading usage probability is compared with the online recommendation threshold, so that high-quality preloaded content (that is, recommended content generated in real time) can be allocated to a target object with a high usage probability. In this way, timeliness and content recommendation quality of content recommended to the target object are ensured, thereby improving attractiveness of the recommended content to the target object, encouraging the target object to participate in content creation and interaction, improving satisfaction and stickiness of the target object with respect to a recommendation platform, improving utilization of the online recommended content, increasing content diversity, and enabling the target object to access real-time high-quality content. In addition, content is recommended in real time based on the first preloading usage probability of the target object, so that computing resources, storage resources, and network bandwidth can be utilized more properly, thereby improving resource utilization.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] To describe the technical solutions in embodiments of this application or in the related art more clearly, the accompanying drawings required for describing embodiments or related art are briefly described below. Apparently, the accompanying drawings in the following descriptions show merely some embodiments of this application, and a person of ordinary skill in the art may still obtain other drawings from these accompanying drawings without creative efforts.

[0018] FIG. 1 is a schematic diagram of an architecture of a data processing system according to an embodiment of this application.

[0019] FIG. 2 is a schematic diagram of a first data processing scenario according to an embodiment of this application.

[0020] FIG. 3 is a first schematic flowchart of a data processing method according to an embodiment of this application.

[0021] FIG. 4 is a schematic diagram of a second data processing scenario according to an embodiment of this application.

[0022] FIG. 5 is a schematic diagram of a third data processing scenario according to an embodiment of this application.

[0023] FIG. 6 is a second schematic flowchart of a data processing method according to an embodiment of this application.

[0024] FIG. 7 is a third schematic flowchart of a data processing method according to an embodiment of this application.

[0025] FIG. 8 is an example stacked line diagram of a preloaded online recommended content quota according to an embodiment of this application.

[0026] FIG. 9 is a fourth schematic flowchart of a data processing method according to an embodiment of this application.

[0027] FIG. 10 is an example diagram of classification of loaded content according to an embodiment of this application.

[0028] FIG. 11 is a schematic diagram of a usage probability of online recommended content under random allocation according to an embodiment of this application.

[0029] FIG. 12 is a schematic diagram of a usage probability of online recommended content under probability-based allocation according to an embodiment of this application.

[0030] FIG. 13 is a schematic diagram of a structure of a data processing apparatus according to an embodiment of this application.

[0031] FIG. 14 is a schematic diagram of a structure of a computer device according to an embodiment of this application.

[0032] The foregoing "first" and "second" are only configured for distinguishing different solutions, and do not represent superiorities of the solutions or priorities in an implementation process.

DETAILED DESCRIPTION

[0033] The technical solutions in embodiments of this application are clearly and completely described below with reference to the accompanying drawings in embodiments of this application. Apparently, the described embodiments are merely some rather than all of embodiments of this application. Based on embodiments of this application, all other embodiments obtained by a person of ordinary skill in the art without involving an inventive effort shall fall within the scope of protection of this application.

[0034] For ease of understanding, some of the terms are first briefly explained below.

[0035] (1) Preloaded content is video content that is transmitted in advance, before a user actively requests playback of a video, from a server to a local cache of a terminal device by using idle network bandwidth and a processing capability of the terminal device.

[0036] (2) Associative storage is a process of binding two or more data items (such as "a first preloading usage probability" and "a device identifier") based on a logical relationship and persistently storing the bound data items. A core objective of the associative storage is to establish searchable associations between data, to facilitate subsequent querying, analysis, and application.

[0037] FIG. 1 is a schematic diagram of an architecture of a data processing system according to an embodiment of this application. As shown in FIG. 1, the system may include a service server 100 and a terminal device cluster. The terminal device cluster may include: a terminal device 200a, a terminal device 200b, a terminal device 200c, …, and a terminal device 200n. The foregoing system may include one or more terminal devices, and a quantity of terminal devices is not limited in this application.

[0038] There is a communication connection between the terminal devices in the cluster. For example, there is a communication connection between the terminal device 200a and the terminal device 200b, and there is a communication connection between the terminal device 200a and the terminal device 200c. In addition, there may be a communication connection between any terminal device in the terminal device cluster and the service server 100. For example, there is a communication connection between the terminal device 200a and the service server 100. A connection manner for the communication connection is not limited. Wired communication may be

used for direct or indirect connection, wireless communication may be used for direct or indirect connection, or another manner may be used. This is not limited in this application.

[0039] An application client may be installed on each terminal device in the terminal device cluster shown in FIG. 1, and when the application client is run on each terminal device, the application client may perform data exchange with the service server 100 shown in FIG. 1, that is, the foregoing communication connection. The application client may be an application client having a content preloading function, such as a video application, a social application, an instant messaging application, a game application, a music application, a shopping application, a novel application, or a browser. The application client may be an independent client, or may be an embedded sub-client integrated in a specific client (for example, a social client, an education client, or a multimedia client). This is not limited herein.

[0040] Using a video application as an example, the service server 100 may be a set of a plurality of servers including a backend server, a data processing server, and the like corresponding to the video application. Therefore, each terminal device may perform data transmission with the service server 100 through an application client corresponding to the video application. For example, the terminal device 200a may send a preloading request to the service server 100 through the application client of the video application, and the service server 100 may generate a preloaded content list for the terminal device 200a based on the preloading request and return the preloaded content list to the terminal device 200a.

[0041] Related data such as user information (for example, a preloading request and a pre-recommended content pool) is involved in a specific implementation of this application. When embodiments of this application are applied to a specific product or technology, user permission or consent is required, and collection, use, and processing of the related data need to comply with related laws, regulations and standards in related countries and regions.

[0042] For ease of subsequent understanding and description, in this embodiment of this application, one terminal device may be selected from the terminal device cluster shown in FIG. 1 for description. For example, the terminal device 200a is used as an example for description. When starting the application client and before outputting content, the terminal device 200a may generate a preloading request in the application client and send the preloading request to the service server 100 through the application client. To achieve high-quality experience of an information stream (a feed stream, also referred to as a content stream), the application client may perform content preloading after being started and before displaying a content output page, in other words, content is preloaded before a target object corresponding to the terminal device 200a browses the content. In

this way, when the target object browses the content, stuttering or slow loading does not occur in the application client, thereby improving viewing experience of the target object.

[0043] Preloading, that is, loading in advance, means that before a target object corresponding to a terminal device actually triggers a network request to pull content from a backend, the content has already been prepared, so that when the target object actually triggers loading, the preloaded content is directly displayed to the target object, thereby providing the target object with an instant loading effect.

[0044] It can be learned from the above that the preloading request in this embodiment of this application is generated before the content is outputted, and is configured for requesting the service server 100 to load the content in advance. A type of the content is not limited in embodiments of this application, and may be set based on an actual application scenario. The content includes, but is not limited to, text content (for example, novel or news), image content, video content, and audio content.

[0045] The service server 100 obtains the preloading request sent by the terminal device 200a, and obtains, based on the preloading request, a first preloading usage probability corresponding to the terminal device 200a, the first preloading usage probability indicating a probability that a target object corresponding to the terminal device 200a interacts with content provided by using a preloading function. The preloading function is configured for responding to the preloading request. The first preloading usage probability may indicate an actual probability that the target object interacts with historical content provided by using the preloading function, or may be used as a predicted probability that the target object interacts with the preloaded content. A process of determining the first preloading usage probability is not described in detail herein. For details, refer to descriptions in embodiments respectively corresponding to FIG. 3 and FIG. 6 below. The preloaded content list may be provided for the terminal device 200a by using the preloading function.

[0046] If the first preloading usage probability is less than an online recommendation threshold, the service server 100 obtains a pre-recommended content pool for the terminal device 200a. A generation timestamp of the pre-recommended content pool is earlier than a generation timestamp of the preloading request. In other words, the pre-recommended content pool is generated in advance. A process of generating the pre-recommended content pool is not described in detail herein. For details, refer to descriptions in an embodiment corresponding to FIG. 9 below. In some embodiments, the service server 100 may generate a target preloaded content list by using the pre-recommended content pool for the terminal device. Because the pre-recommended content pool

is generated in advance, the service server 100 may avoid recommending online content in real time for the terminal device 200a, and instead use pre-computed content as preloaded content, thereby saving computing resources for the online content (because the first preloading usage probability is less than the online recommendation threshold).

[0047] If the first preloading usage probability is greater than or equal to the online recommendation threshold, the service server 100 generates an online recommended content set for the terminal device. A generation timestamp of the online recommended content set is later than the generation timestamp of the preloading request. In other words, the service server recommends online content in real time, and uses the online content as the preloaded content. The service server 100 generates a target preloaded content list based on the online recommended content set. Compared with pre-recommended content (that is, online content not recommended in real time), the online content recommended in real time has high quality, in other words, better matches a browsing characteristic of the target object, to increase a probability that the target object interacts with the preloaded content, thereby improving utilization, that is, an output rate, of the preloaded content (where the preloaded content is online recommended content in this case).

[0048] It can be learned from the above that this embodiment of this application provides a resource allocation method based on the first preloading usage probability. For a terminal device corresponding to the first preloading usage probability that is greater than or equal to the online recommendation threshold, the online recommended content set generated in real time is provided, and the preloaded content list generated in real time can further be provided. For a terminal device corresponding to the first preloading usage probability that is less than the online recommendation threshold, the preloaded content is allocated from the pre-recommended content pool (that is, a recommended content pool generated in advance). Therefore, resource costs for real-time computation can be reduced by using the pre-recommended content pool. The first preloading usage probability is compared with the online recommendation threshold, so that high-quality preloaded content (that is, recommended content generated in real time) can be allocated to a target object with a high usage probability, thereby improving content recommendation quality and improving utilization of the online recommended content.

[0049] The service server 100, the terminal device 200a, the terminal device 200b, the terminal device 200c, ..., and the terminal device 200n may all be blockchain nodes in a blockchain network. Data (for example, the preloading request and the pre-recommended content pool) described in this application may be stored in a manner in which a blockchain node generates a block based on the data and adds the block into a blockchain for storage.

[0050] The blockchain is a new-type application mode of a computer technology such as distributed data storage, peer-to-peer transmission, a consensus mechanism, and an encryption algorithm, and is mainly configured to organize data in chronological order and encrypt the data into a ledger, so that the data cannot be tampered with and forged, and data verification, storage, and updating can be performed. The blockchain is essentially a decentralized database. Each node in the database stores an identical blockchain. The blockchain network may classify nodes into core nodes, data nodes, and light nodes. The core nodes, the data nodes, and the light nodes are collectively referred to as blockchain nodes.

[0051] The core nodes are responsible for consensus of the whole blockchain network. In other words, the core nodes are consensus nodes in the blockchain network. A process of writing transaction data into a ledger in the blockchain network may be as follows. A data node or a light node in the blockchain network obtains the transaction data, and transmits the transaction data across the blockchain network (that is, the node transmits the transaction data in a relay manner) until a consensus node receives the transaction data. The consensus node then packages the transaction data into a block, performs consensus on the block, and writes the transaction data to the ledger after the consensus is completed. The preloading request and the pre-recommended content pool are used as examples of the transaction data. After consensus on the transaction data is completed, the service server 100 (a blockchain node) generates a block based on the transaction data, and stores the block in the blockchain network. For reading of the transaction data (that is, the preloading request and the pre-recommended content pool), a blockchain node may obtain, from the blockchain network, the block including the transaction data, and obtain the transaction data from the block.

[0052] The method according to this embodiment of this application may be performed by a computer device. The computer device includes, but is not limited to, a terminal device or a service server. The service server may be an independent physical server, or may be a server cluster or distributed system including a plurality of physical servers, or may be a cloud server providing basic cloud computing services such as a cloud database, a cloud service, cloud computing, a cloud function, cloud storage, a network service, cloud communication, a middleware service, a domain name service, a security service, a content delivery network (CDN), a big data and artificial intelligence platform. The terminal device includes, but is not limited to, a mobile phone, a computer, an intelligent voice interaction device, a smart home appliance, an on-board terminal, and an aircraft. The terminal device may be directly or indirectly connected to the service server in a wired or wireless manner. This is not limited in embodiments of this application.

[0053] FIG. 2 is a schematic diagram of a first data processing scenario according to an embodiment of this application. This embodiment of this application may be applied to various scenarios, including but not limited to a cloud technology, artificial intelligence, intelligent transportation, assisted driving, and audio and video. This embodiment of this application may be applied to service scenarios such as a multimedia content query scenario, a multimedia content recommendation scenario, and a multimedia content distribution scenario. Specific service scenarios are not enumerated herein one by one. An implementation process of the data processing scenario may be performed by a service server, or may be performed through interaction between a terminal device and the service server. This is not limited herein. For ease of description and understanding, an example in which the implementation process is performed through interaction between a terminal device and a service server is used for description in this embodiment of this application. In other words, the terminal device and the service server are used as computer devices. The terminal device may be any terminal device in the terminal device cluster in this embodiment corresponding to FIG. 1, and the service server may be the service server 100 in the embodiment corresponding to FIG. 1.

[0054] As shown in FIG. 2, there is a binding relationship between a target object 20a and a terminal device 20b, and the binding relationship may be a relationship in which the target object 20a is a user of the terminal device 20b. An application client is installed on the terminal device 20b. For ease of understanding, a multimedia application is used as an example of the application client in FIG. 2. The terminal device 20b may display a non-content output page 201b and system time (that is, current time) in the multimedia application. The non-content output page 201b does not include a content detail page (or may be referred to as an underlying content page). For example, if content is video content, the non-content output page 201b does not include a video playback page, in other words, the non-content output page 201b may include a page in the multimedia application other than the video playback page. For example, if the content is text content, the non-content output page 201b does not include a text detail page. In other words, the non-content output page 201b may include a page in the multimedia application other than the text detail page. A type of the content is not limited in embodiments of this application, and the content may include one or more of the following: text content, video content, image content, and audio content.

[0055] In this embodiment of this application, for ease of understanding, video content is used as an example of the content, in other words, content requested in the preloading request is the video content. Refer to FIG. 2. When the non-content output page 201b is displayed, in other words, before a content (for example, video, novel, or information) playback page is entered, the terminal device 20b generates a preloading request 20c at 12:00 on January 1, 2024. The preloading request

20c may include a device identifier indicating the terminal device 20b, for example, an internet protocol address (IP address for short), or an application account logged in to the multimedia application. A type of the device identifier is not limited in embodiments of this application, and may be set based on an actual application scenario. For ease of description, a device 1 is used as an example of the device identifier in FIG. 2.

[0056] In some embodiments, the terminal device 20b sends the preloading request 20c to a service server 100. After obtaining the preloading request 20c sent by the terminal device 20b, the service server 100 may obtain an index table 20d based on the preloading request 20c. The index table 20d may include an index key based on a device identifier and an index value based on a preloading usage probability (which is briefly referred to as a probability in FIG. 2). FIG. 2 shows three examples of index key-value pairs in the index table 20d. A first index key-value pair includes an index key generated based on the device 1 and an index value generated based on 0.4. The first index key-value pair indicates that a current preloading usage probability of a terminal device indicated by the device 1 is 0.4. A second index key-value pair includes an index key generated based on a device 2 and an index value generated based on 0.3. The second index key-value pair indicates that a current preloading usage probability of a terminal device indicated by the device 2 is 0.3. A third index key-value pair includes an index key generated based on a device 3 and an index value generated based on 0.5. The third index key-value pair indicates that a current preloading usage probability of a terminal device indicated by the device 3 is 0.5.

[0057] The foregoing descriptions refer to a current preloading usage probability instead of a first preloading usage probability because a preloading usage probability is periodically updated. In other words, the service server 100 periodically updates a preloading usage probability corresponding to each terminal device. Specifically, the service server 100 updates a preloading usage probability of a terminal device based on a first update period (for example, half an hour or one hour). For details, refer to descriptions in embodiments respectively corresponding to FIG. 3 and FIG. 6 below.

[0058] As shown in FIG. 2, the service server 100 may search the index table 20d, and determine, as a first preloading usage probability 20e corresponding to the terminal device 20b, an index value (that is, 0.4) that has an index relationship with the device 1 indicating the terminal device 20b. The first preloading usage probability 20e indicates a probability that the target object 20a interacts with content provided by using a preloading function. The content provided by using the preloading function may include historical content, that is, content pulled within a historical time period (which is earlier than 12:00 on January 1, 2024 shown in FIG. 2). In this case, the first

preloading usage probability 20e may be considered an actual probability that the target object 20a interacts with the historical content. The content provided by the preloading function may alternatively include content to be pulled (that is, preloaded content). In this case, the first preloading usage probability 20e may be considered a predicted probability that the target object 20a interacts with the preloaded content (to-be-generated content).

[0059] In some embodiments, the service server 100 compares the first preloading usage probability 20e with an online recommendation threshold. If the first preloading usage probability 20e (which is 0.4 shown in FIG. 2) is less than the online recommendation threshold, for example, the online recommendation threshold is equal to 0.45, the service server 100 may determine that a probability that the terminal device 20b outputs the preloaded content is low, that is, a probability that the target object 20a triggers the preloaded content is low. In other words, a probability that the target object 20a views the preloaded content is low. In this case, the service server 100 obtains a pre-recommended content pool 20h corresponding to the terminal device 20b. The pre-recommended content pool 20h has been generated in advance, to be specific, a generation timestamp 201h of the pre-recommended content pool 20h is earlier than a generation timestamp of the preloading request 20c. As shown in the example in FIG. 2, the generation timestamp 201h of the pre-recommended content pool 20h is 00:01 on January 1, 2024, and the generation timestamp of the preloading request 20c is 12:00 on January 1, 2024. A process of generating the pre-recommended content pool 20h is not described in detail in this embodiment of this application. For details, refer to descriptions in an embodiment corresponding to FIG. 9 below.

[0060] The pre-recommended content pool 20h shown in the example in FIG. 2 includes a video 1, a video 2, a video 3, and a video 4. The service server 100 may generate a target preloaded content list 20i based on the pre-recommended content pool 20h. The preloaded content list 20i shown in the example in FIG. 2 includes the video 1, the video 2, and the video 3. An amount of preloaded content in the preloaded content list 20i is not limited in this embodiment of this application and may be set based on an actual application scenario. In some embodiments, the service server 100 may return the preloaded content list 20i to the terminal device 20b. Subsequently, if the target object 20a triggers content output, for example, triggers video playback, the terminal device 20b does not need to initiate a content obtaining request to the service server 100, and may output, in sequence, preloaded content, for example, the video 1, the video 2, and the video 3 shown in FIG. 2, in the preloaded content list 20i.

[0061] The foregoing describes a scenario in which the first preloading usage probability 20e is less than the online recommendation threshold. If the first preloading usage probability 20e

(which is 0.4 shown in FIG. 2) is greater than or equal to the online recommendation threshold, for example, the online recommendation threshold is equal to 0.36, the service server 100 may determine that the probability that the terminal device 20b outputs the preloaded content is high, that is, the probability that the target object 20a triggers the preloaded content is high. In other words, the probability that the target object 20a views the preloaded content is high. In this case, the service server 100 generates an online recommended content set 20f for the terminal device 20b. Different from the pre-recommended content pool 20h, a generation timestamp 201f of the online recommended content set 20f is later than the generation timestamp of the preloading request 20c. As shown in the example in FIG. 2, the generation timestamp 201f is 12:01 on January 1, 2024. To be specific, the service server 100 generates the online recommended content set 20f in real time for the preloading request 20c sent by the terminal device 20b. A process of generating the online recommended content set 20f is not described in detail in this embodiment of this application. For details, refer to descriptions in an embodiment corresponding to FIG. 3 below.

[0062] In the example in FIG. 2, the online recommended content set 20f includes a video a, a video b, a video c, and a video d. The service server 100 may generate a target preloaded content list 20g based on the online recommended content set 20f. As shown in the example in FIG. 2, the target preloaded content list 20g may include the video a, the video b, and the video c. The target preloaded content list 20g and the target preloaded content list 20i have the same function. Therefore, a subsequent processing process of the target preloaded content list 20g is not described herein again. For details, refer to the foregoing descriptions of the target preloaded content list 20i.

[0063] It can be learned from the above that this embodiment of this application provides a resource allocation method based on the first preloading usage probability. For a terminal device corresponding to the first preloading usage probability that is greater than or equal to the online recommendation threshold, the online recommended content set generated in real time is provided, and the preloaded content list generated in real time can further be provided. For a terminal device corresponding to the first preloading usage probability that is less than the online recommendation threshold, the preloaded content is allocated from the pre-recommended content pool (that is, a recommended content pool generated in advance). Therefore, resource costs for real-time computation can be reduced by using the pre-recommended content pool. The first preloading usage probability is compared with the online recommendation threshold, so that high-quality preloaded content (that is, recommended content generated in real time) can be allocated to a target object with a high usage probability, thereby improving content recommendation quality and improving utilization of the online recommended content.

[0064] FIG. 3 is a first schematic flowchart of a data processing method according to an embodiment of this application. An implementation process of the data processing method may be performed by a service server, or may be performed through interaction between a terminal device and the service server. This is not limited herein. For ease of description and understanding, an example in which the implementation process is performed by a service server is used for description in this embodiment of this application. In other words, the service server is used as a computer device. The service server may be the service server 100 in the embodiment corresponding to FIG. 1. As shown in FIG. 3, the data processing method may include at least the following operation S101 to operation S104.

[0065] Operation S101: Obtain a preloading request sent by a terminal device, and obtain, based on the preloading request, a first preloading usage probability corresponding to the terminal device,

[0066] the first preloading usage probability indicating a probability that a target object corresponding to the terminal device interacts with preloaded content. The preloaded content may be content provided by using a preloading function, and the preloading function is a function started in response to the preloading request.

[0067] In some embodiments, the preloading request sent by the terminal device is obtained at a first moment. In this case, when the first preloading usage probability corresponding to the terminal device is determined, a quantity of application startup times and a quantity of preloading usage times of the terminal device within a first time period is determined. The latest timestamp of the first time period is earlier than the first moment. A quantity ratio between the quantity of application startup times and the quantity of preloading usage times is determined, and the first preloading usage probability is determined based on the quantity ratio. In some embodiments, the first preloading usage probability and a device identifier indicating a terminal device may be associatively stored in a database.

[0068] In some embodiments, an obtaining timestamp of the preloading request is determined, an online recommended content set for the terminal device is generated if the obtaining timestamp is within a first-type time period, and a target preloaded content list is generated based on the online recommended content set. The first preloading usage probability corresponding to the terminal device is obtained if the obtaining timestamp is within a second-type time period. A total amount of output content corresponding to the second-type time period is greater than a total amount of output content corresponding to the first-type time period. The first-type time period may

be a content output off-peak time period. For example, the first-type time period may be from 1:00 a.m. to 7:00 a.m. every day.

[0069] A preloading usage probability corresponding to the terminal device is time-sensitive. In this embodiment of this application, the preloading usage probability corresponding to the terminal device within a current time period is referred to as the first preloading usage probability. The service server sets, to a first update period T1, an update period of the preloading usage probability corresponding to the terminal device. In this case, the service server updates the preloading usage probability corresponding to the terminal device every T1 hours. Assuming that T1 is 0.5, the service server updates, every half an hour, the preloading usage probability corresponding to the terminal device, that is, generates a new preloading usage probability. Assuming that a current preloading usage probability (which may be understood as the first preloading usage probability) of the terminal device is generated by the service server at 12:00 on May 2, 2024, the service server regenerates the preloading usage probability of the terminal device at 12:30 on May 2, 2024. In this embodiment of this application, a preloading usage probability corresponding to the next update period of the current preloading usage probability is referred to as a second preloading usage probability.

[0070] The data processing method provided in this embodiment of this application may be applied to a plurality of types of multimedia data and product applications, including but not limited to short videos, long videos, and image-text content. Each application focuses on a different data dimension. For example, an application 1 focuses on duration for which a target object browses a content playback page (for example, a video playback page), an application 2 focuses on a quantity of times that the target object triggers the content playback page, and an application 3 focuses on a content type of content outputted on the content playback page, for example, whether the output content is online recommended data or pre-recommended data. Because each application focuses on a different data dimension, the method used by the service server to determine the preloading usage probability corresponding to the terminal device is not limited in embodiments of this application, and may be set based on an actual application scenario.

[0071] An embodiment of this application describes a simple method for determining a preloading usage probability. When updating the preloading usage probability, the service server collects statistics on a quantity of application startup times and a quantity of preloading usage times of the terminal device within a first time period. The first time period is a time period corresponding to the foregoing first update period. For example, if the first update period is half an hour, duration corresponding to the first time period is half an hour. For example, if the service server generates

the preloading usage probability corresponding to the terminal device at 12:00 and generates a new preloading usage probability corresponding to the terminal device at 12:30, the first time period may include 12:00 to 12:30. The quantity of application startup times of the terminal device is a quantity of times that an application client is started, and the quantity of preloading usage times is a quantity of times that the terminal device enters a content output page. In other words, the application client is not only started, but also enters the content output page.

[0072] In some embodiments, the service server determines a quantity ratio between the quantity of application startup times and the quantity of preloading usage times, and may determine the quantity ratio as the first preloading usage probability, or further determine the first preloading usage probability in combination with other dimensional data of the application client. With reference to the foregoing example, if a generation timestamp of the first preloading usage probability is 12:30, the preloading usage probability corresponding to the terminal device is the first preloading usage probability within a time period from 12:30 to 13:00, and the service server regenerates a new preloading usage probability at 13:00. The service server associatively stores, in the database, for example, in the index table 20d shown in FIG. 2, the first preloading usage probability and the device identifier indicating the terminal device.

[0073] Refer also to FIG. 4. FIG. 4 is a schematic diagram of a second data processing scenario according to an embodiment of this application. As shown in FIG. 4, a service server 100 obtains a preloading request 20c. The preloading request 20c carries a device identifier indicating a terminal device. A device 1 is used as an example of the device identifier in FIG. 4. The service server 100 determines an obtaining timestamp of the preloading request 20c, that is, a timestamp at which the preloading request 20c is obtained. If the obtaining timestamp is within a first-type time period 30a, where the first-type time period 30a shown in FIG. 4 is from 1:00 to 9:00 in a day, for example, if the obtaining timestamp is 2:00 a.m., the service server 100 generates an online recommended content set 20f for the terminal device. For a process of generating the online recommended content set 20f, refer to descriptions in operation S103 below. Details are not described herein.

[0074] The service server 100 generates a target preloaded content list based on the online recommended content set 20f. A specific process in which the service server generates the target preloaded content list based on the online recommended content set 20f is the same as a specific process in which the service server generates the target preloaded content list based on a pre-recommended content pool. Therefore, for details, refer to the specific process of generating the preloaded content list in operation S102 below.

[0075] If the obtaining timestamp is within a second-type time period 30b, where the second-type time period 30b shown in FIG. 4 is from 10:11 a.m. on a current day to 1:00 a.m. on the next day, for example, if the obtaining timestamp is 8:00 p.m., the service server 100 obtains a first preloading usage probability 20e corresponding to the terminal device. In FIG. 4, the first preloading usage probability 20e is shown as 0.4. For output content of a full set of terminal devices, a total amount of output content corresponding to the second-type time period is greater than a total amount of output content corresponding to the first-type time period. Therefore, the first-type time period 30a may be understood as a content output off-peak time period, that is, a time period in which fewer target objects view content (for example, video or novel) through the terminal device. The second-type time period 30b may be understood as a content output peak time period, that is, a time period within which more target objects view content through the terminal device.

[0076] Because a total amount of output content corresponding to the content output off-peak time period is small, system resources of the service server are in an idle state, in other words, utilization of the system resources is low. To maximize utilization of the system resources, high-quality online recommended content is provided to the terminal device within the content output off-peak time period, to be specific, the online recommended content is generated in real time, and the online recommended content is pushed to the terminal device as preloaded content. It can be learned from the above that, according to this embodiment of this application, backend system resources can be properly used, and high-quality preloaded content (generated from online recommended content) can be provided to the target object.

[0077] Correspondingly, because a total amount of output content corresponding to the content output peak time period is large, in other words, a large quantity of target objects view content (for example, video, information, or novel) online through the terminal device, system resources of the service server are in a busy state. In this case, utilization of the system resources is high. To properly allocate the system resources, whether the target object plays preloaded content needs to be determined. Therefore, the service server obtains the first preloading usage probability, and performs further determining based on the first preloading usage probability. For details, refer to the following operation S102 and operation S103.

[0078] Operation S102: Determine whether the first preloading usage probability is less than an online recommendation threshold.

[0079] The online recommendation threshold is a probability threshold for performing online content recommendation. Determining of the online recommendation threshold is associated with current preloading usage probabilities respectively corresponding to the full set of terminal

devices, and is also associated with a real-time second online recommendation quota. A process of generating the online recommendation threshold is not described in detail herein. For details, refer to descriptions in an embodiment corresponding to FIG. 7 below. The service server compares the first preloading usage probability corresponding to the terminal device with the online recommendation threshold. If the first preloading usage probability is less than the online recommendation threshold, operation S103 is performed; or if the first preloading usage probability is greater than or equal to the online recommendation threshold, operation S104 is performed.

[0080] Operation S103: Generate a target preloaded content list based on a pre-recommended content pool for the terminal device, a generation timestamp of the pre-recommended content pool being earlier than a generation timestamp of the preloading request.

[0081] In some embodiments, F pieces of pre-recommended content that are in a valid state are determined from the pre-recommended content pool for the terminal device. F is a positive integer. Recommendation scores respectively corresponding to the F pieces of pre-recommended content are determined, and G recommendation scores are obtained from the F recommendation scores. The G recommendation scores are greater than or equal to H recommendation scores. The H recommendation scores include a recommendation score in the F recommendation scores other than the G recommendation scores. G is a positive integer less than or equal to F. H is a positive integer less than F. G pieces of pre-recommended content respectively corresponding to the G recommendation scores are obtained from the F pieces of pre-recommended content, and the target preloaded content list is generated based on the G pieces of pre-recommended content.

[0082] A process of generating the target preloaded content list based on the G pieces of pre-recommended content may include: generating a first preloaded content list based on the G pieces of pre-recommended content, the G pieces of pre-recommended content including first pre-recommended content and second pre-recommended content; determining a content similarity between the first pre-recommended content and the second pre-recommended content; deleting the first pre-recommended content from the first preloaded content list if the content similarity is greater than or equal to a similarity threshold, to obtain a second preloaded content list; obtaining a largest recommendation score from the H recommendation scores, and obtaining, from the F pieces of pre-recommended content, third pre-recommended content corresponding to the largest recommendation score; adding the third pre-recommended content to the second preloaded content list, to obtain a third preloaded content list; and generating the target preloaded content list based on the third preloaded content list, a content similarity between any two pieces of preloaded content in the target preloaded content list being less than the similarity threshold.

[0083] The first pre-recommended content and the second pre-recommended content are any two pieces of pre-recommended content in the G pieces of pre-recommended content. In other words, similarity calculation needs to be performed on any two pieces of pre-recommended content in the G pieces of pre-recommended content, to determine a content similarity between the two pieces of pre-recommended content. For ease of understanding, video is used as an example of the preloaded content in this embodiment of this application. For example, G is 4, and the four pieces of pre-recommended content are a video 1, a video 2, a video 3, and a video 4. In this case, content similarities between the video 1 and the video 2, between the video 2 and the video 3, between the video 3 and the video 4, between the video 1 and the video 3, between the video 1 and the video 4, and between the video 2 and the video 4 need to be calculated.

[0084] In some embodiments, when the content similarity between the first pre-recommended content and the second pre-recommended content is determined, multi-modal information such as visual information, audio information, and text information needs to be comprehensively considered. During implementation, a first key frame may be determined from the first pre-recommended content, and a second key frame may be determined from the second pre-recommended content. Image feature vectors of the first key frame and the second key frame are extracted, and an image similarity between the first key frame and the second key frame is determined based on the two image feature vectors. When there are a plurality of key frames, pairwise comparison (for example, cosine similarity comparison) needs to be performed on all key frame features of the first pre-recommended content and the second pre-recommended content to obtain an average value or a maximum value. Then, to determine an audio similarity, a first audio feature of the first pre-recommended content and a second audio feature of the second pre-recommended content may be extracted, and the audio similarity is determined based on the first audio feature and the second audio feature. To determine a text similarity, word embedding processing may be performed on first text information in the first pre-recommended content to obtain a first text vector, and word embedding processing may be performed on second text information in the second pre-recommended content to obtain a second text vector. The text similarity between the first pre-recommended content and the second pre-recommended content is determined based on the first text vector and the second text vector. Finally, weighting processing is performed on the image similarity, the audio similarity, and the text similarity to obtain the content similarity between the first pre-recommended content and the second pre-recommended content.

[0085] After obtaining the preloading request, the service server calculates all candidate content in real time, to determine the online recommended content. In this case, a complete recommendation process needs to be performed, resulting in high resource consumption. In addition,

because preloading needs to be initiated before the target object browses content (for example, video, novel, or information), but not all content requested by preloading requests is used by the target object, a large quantity of unnecessary requests are added, resulting in huge request and cost pressures on backend services. Based on this, in this embodiment of this application, a terminal device that initiates a preloading request is determined, to determine which type of preloaded content is returned. The preloaded content may include two types. A first type is pre-recommended content, which is obtained from a pre-recommended content pool generated in advance, and a second type is online recommended content, which is obtained from an online recommended content set generated in real time.

[0086] Refer to FIG. 4. If the first preloading usage probability 20e is less than the online recommendation threshold, for example, if the online recommendation threshold is 0.45, the service server 100 obtains a pre-recommended content pool 20h corresponding to the terminal device, and generates a target preloaded content list based on the pre-recommended content pool 20h for the terminal device.

[0087] The following describes a process of generating a target preloaded content list based on a pre-recommended content pool for the terminal device. For ease of understanding, video is used as an example of preloaded content in this embodiment of this application. Refer also to FIG. 5. FIG. 5 is a schematic diagram of a third data processing scenario according to an embodiment of this application. A generation timestamp 30e of a pre-recommended content pool 30f corresponding to a terminal device shown in FIG. 5 is at 00:01 on January 1, 2024. The pre-recommended content pool 30f includes a video 1, a video 2, a video 3, a video 4, a video 5, a video 6, a video 7, and the like. A service server determines, from the pre-recommended content pool 30f, a video in an invalid state, and the video is referred to as an invalid video for short. A reason why the video is in the invalid state is not limited in this embodiment of this application, and may be set based on an actual application scenario. The reason includes, but is not limited to, "has been played", "deleted by a publisher", "the publisher has been blocked", or the like, as shown in FIG. 5.

[0088] As shown in the example in FIG. 5, because the video 1 has been played, the video 1 is an invalid video. For example, the service server determines that an obtaining timestamp of a currently received preloading request is 12:00 on January 1, 2024. If the terminal device has initiated a historical preloading request during a period from 00:01 on January 1, 2024 to 12:00 on January 1, 2024, for example, the terminal device sent a preloading request to the service server at 11:00 on January 1, 2024, the preloading request may be referred to as a historical preloading request relative to system time (that is, current time). In response to the historical preloading request,

the service server returns the video 1 in the pre-recommended content pool 30f to the terminal device as preloaded content, in other words, generates a preloaded content list including the video 1, and returns the preloaded content list to the terminal device. The terminal device obtains the preloaded content list including the video 1. When a target object triggers content playback, the video 1 is played. Therefore, the video 1 is a played video, and the played video is an invalid video.

[0089] As shown in the example in FIG. 5, because a publisher corresponding to the video 5 has deleted the video 5, a service platform does not have permission to play the video 5. In this case, the video 5 is an invalid video. Because the target object has blocked a publisher (for example, a publisher 300) of the video 6, in other words, has set not to watch a video published by the publisher 300, the video 6 is an invalid video.

[0090] FIG. 5 shows three types of invalid videos. In actual application, other invalid types may also be included. In addition, the preloaded content shown in FIG. 5 is video. If the preloaded content is audio or text, the invalid types may include, but are not limited to, having been outputted, having been deleted, having been blocked, and the like.

[0091] The service server deletes videos (that is, pre-recommended content) that are in an invalid state from the pre-recommended content pool 30f to obtain pre-recommended content that is in a valid state. For example, pre-recommended content 30g shown in FIG. 5 includes the video 2, the video 3, the video 4, and the video 7. In other words, F shown in FIG. 5 is 4. In some embodiments, the service server obtains a recommendation score corresponding to each piece of pre-recommended content. For example, in FIG. 5, a recommendation score of the video 2 is 0.87, a recommendation score of the video 3 is 0.85, a recommendation score of the video 4 is 0.83, and a recommendation score of the video 7 is 0.8. For ease of understanding, G shown in FIG. 5 is 3. According to a descending order of recommendation scores, the service server obtains, from the pre-recommended content 30g, pre-recommended content whose recommendation scores rank in the top three, that is, the video 2, the video 3, and the video 4.

[0092] In some embodiments, the service server may generate a first preloaded content list based on the video 2, the video 3, and the video 4, and obtain a content similarity between any two pieces of pre-recommended content in the first preloaded content list. For example, in FIG. 5, a similarity between the video 2 and the video 3 is 0.7, a similarity between the video 2 and the video 4 is 0.3, and a similarity between the video 3 and the video 4 is 0.2. The video 2, the video 3, or the video 4 may be the foregoing first pre-recommended content, and the second pre-recommended content may be one of the remaining two videos. The service server separately compares each content similarity with a similarity threshold. If the three similarities shown in FIG. 5 are all less

than the similarity threshold, for example, if the similarity threshold is 0.8, the service server may determine the first preloaded content list as a preloaded content list 30i, which includes the video 2, the video 3 and the video 4; and return the preloaded content list 30i to the terminal device.

[0093] If a content similarity is greater than or equal to the similarity threshold, for example, if the similarity threshold is 0.5, the similarity 0.7 between the video 2 and the video 3 shown in FIG. 5 is greater than 0.5. The service server deletes, from the two videos, a video with a smaller recommendation score, that is, deletes the video 3 from the first preloaded content list, to obtain a second preloaded content list. The second preloaded content list obtained after the video 3 is deleted includes two pieces of pre-recommended content, that is, the video 2 and the video 4. Therefore, the service server obtains, from the pre-recommended content 30g, pre-recommended content that is not included in the second pre-recommended content list and that has a largest recommendation score. The pre-recommended content shown in FIG. 5 is the video 7. The service server adds the video 7 to the second preloaded content list obtained after the video 3 is deleted, to obtain a third preloaded content list 30h.

[0094] A processing process performed by the service server on the third preloaded content list 30h is the same as a processing process performed on a candidate preloaded content list including the video 2, the video 3, and the video 4. Details are not described herein again. If a similarity between two pieces of pre-recommended content in the third preloaded content list 30h is greater than or equal to the similarity threshold, the service server continues to obtain pre-recommended content from the pre-recommended content 30g. If the pre-recommended content 30g does not include any pre-recommended content other than the third preloaded content list 30h, the service server may determine the third preloaded content list 30h as a target preloaded content list.

[0095] A feasible implementation is as follows: When obtaining the preloading request sent by the terminal device, if the service server determines that a preloading usage probability corresponding to the terminal device is less than an online recommendation threshold, and that a total amount of pre-recommended content included in a pre-recommended content pool corresponding to the terminal device is less than G, the service server may generate an online recommended content set for the terminal device in real time, and further generate the target preloaded content list based on the online recommended content set.

[0096] The preloaded content list in this embodiment of this application is referred to as an information stream, that is, a feed stream. The feed stream is an information stream that is

continuously updated and that presents content to the target object, and can meet a requirement of the target object for specific content through personalized recommendation.

[0097] Operation S104: Generate the online recommended content set for the terminal device, and generate the target preloaded content list based on the online recommended content set, a generation timestamp of the online recommended content set being later than the generation timestamp of the preloading request.

[0098] In some embodiments, a total quantity of preloading usage probabilities that are greater than or equal to the online recommendation threshold is determined. If the total quantity is less than or equal to a first online recommendation quota, the online recommended content set for the terminal device is generated based on an output content feature of the terminal device and a candidate content pool corresponding to system time. If the total quantity is greater than the first online recommendation quota, the target preloaded content list is generated based on the pre-recommended content pool.

[0099] When the first preloading usage probability is greater than or equal to the online recommendation threshold, the service server determines a first online recommendation quota corresponding to a current moment. The first online recommendation quota is a parameter that changes in real time, and is determined based on a second online recommendation quota that changes in real time and an online recommendation total quota that has a fixed attribute. A relationship among the three quotas may be represented by using Formula (1):

[00100] Online recommendation total quota≥Second online recommendation quota+First online recommendation quota (1)

[00101] The online recommendation total quota is preset and fixed, and is determined based on system resources of the service server. The second online recommendation quota is a total amount of non-preloaded online recommended content. The non-preloaded online recommended content is content pulled by the terminal device after the terminal device enters a content playback page. In other words, the second online recommendation quota is a second total amount of content pulled by the terminal device after the terminal device enters the content playback page. Preloaded online recommended content is content pulled by the terminal device before the terminal device enters the content playback page. In other words, the first online recommendation quota is a first total amount of content pulled by the terminal device before the terminal device enters the content playback page. In this embodiment of this application, the first online recommendation quota that changes in real time is determined based on the second online recommendation quota that changes in real time and the fixed online recommendation total quota. Specifically, the service server may

collect an actual second online recommendation quota within a historical time period as a second online recommendation quota at a current moment. For example, an actual second online recommendation quota in the previous second is determined as the second online recommendation quota at the current moment. Therefore, difference calculation may be performed on the online recommendation total quota and the second online recommendation quota in the last second, to obtain the first online recommendation quota at the current moment.

[00102] The service server collects statistics on a total quantity of preloading usage probabilities that are greater than or equal to the online recommendation threshold. A total quantity 30c shown in FIG. 4 is 1.5 million. In other words, 1.5 million or more terminal devices initiate preloading requests to the service server within a current time period (for example, one second or half a minute), and there are preloading usage probabilities (including the foregoing first preloading usage probability) that respectively correspond to 1.5 million terminal devices and that are all greater than or equal to the online recommendation threshold.

[00103] The service server compares the total quantity 30c with the first online recommendation quota. The process is mainly to ensure that the service server is not overloaded and that system resources can be utilized to a maximum extent. Therefore, if the total quantity 30c is less than or equal to the first online recommendation quota, for example, if the first online recommendation quota is 1.51 million, the service server obtains an output content feature of the terminal device and a candidate content pool 30d corresponding to the system time. The output content feature of the terminal device is a feature of content (including preloaded content and non-preloaded content) outputted by the terminal device within a historical time period. A feasible manner of obtaining the output content feature is as follows: The service server inputs content outputted by the terminal device within a historical time period, for example, video outputted within a week, into a video recognition model. In addition, the service server performs, by using the video recognition model, recognition and integration processing on the videos outputted within the week, to obtain features corresponding to the videos outputted within the week. For example, the videos outputted by the terminal device within the week have features of comedy and cute pets. The video recognition model is a trained AI model. A model type and a structure of the video recognition model are not limited in embodiments of this application, and may be set based on a requirement of an actual application scenario.

[00104] A manner of generating the candidate content pool 30d is as follows: The candidate content pool 30d is generated based on candidate content for which the service server has output permission at the current time. To narrow the scope, another manner of generating the candidate

content pool 30d is as follows: The candidate content pool 30d is generated based on candidate content associated with the terminal device. For example, if an application account is bound to the terminal device, content published by a friend account that has a friend relationship with the application account, content outputted by the friend account, and content with which the friend account interacts (for example, comments, likes, rewards, or follows) may all be added to the candidate content pool 30d. In addition, content published by a content publisher with which an application client installed on the terminal device interacts may be added to the candidate content pool 30d.

[00105] The service server identifies a feature of candidate content in the candidate content pool 30d. For example, in FIG. 4, the candidate content pool 30d includes a video 1, a video 2, a video 3, a video 4, a video 5, …, a video a, a video b, a video c, a video d, and a video e. Therefore, the service server may perform recognition processing on the video 1 by using the foregoing video recognition model, to obtain a video feature of the video 1. In some embodiments, a feature similarity 1 between the video feature of the video 1 and the output content feature is identified. If the feature similarity 1 is greater than or equal to a similarity threshold z (where a similarity threshold herein is referred to as the similarity threshold z to distinguish the similarity threshold from the foregoing similarity threshold), the service server may determine the video 1 as online recommended content. If the feature similarity 1 is less than the similarity threshold z, the video 1 is filtered out, and the next video (for example, the video 2) is processed in a manner similar to the processing process of the video 1 shown in FIG. 4, until recognition of all candidate content in the candidate content pool 30d is completed. All online recommended content is determined as an online recommended content set. The online recommended content set 20f shown in FIG. 4 includes the video a, the video b, the video c, the video d, and the like.

[00106] Similar to the pre-recommended content pool 30f shown in FIG. 5, each piece of online recommended content in the online recommended content set also corresponds to a recommendation score. Therefore, the service server may determine a target preloaded content list from the online recommended content set based on the recommendation score of each piece of the online recommended content. A process in which the service server generates the target preloaded content list based on the online recommended content set is the same as the process in which the service server generates the target preloaded content list based on the pre-recommended content pool. Therefore, details are not described herein again. For the details, refer to the descriptions in operation S102 above.

[00107] In another scenario, refer to FIG. 4. If the total quantity 30c is greater than the first online recommendation quota, for example, if the first online recommendation quota is 1.4 million, the service server generates the target preloaded content list based on the pre-recommended content pool 20h.

[00108] It can be learned from the above that this embodiment of this application provides a resource allocation method based on the first preloading usage probability. For a terminal device corresponding to the first preloading usage probability that is greater than or equal to the online recommendation threshold, the online recommended content set generated in real time is provided, and the preloaded content list generated in real time can further be provided. For a terminal device corresponding to the first preloading usage probability that is less than the online recommendation threshold, the preloaded content is allocated from the pre-recommended content pool (that is, a recommended content pool generated in advance). Therefore, resource costs for real-time computation can be reduced by using the pre-recommended content pool. The first preloading usage probability is compared with the online recommendation threshold, so that high-quality preloaded content (that is, recommended content generated in real time) can be allocated to a target object with a high usage probability, thereby improving content recommendation quality and improving utilization of the online recommended content.

[00109] FIG. 6 is a second schematic flowchart of a data processing method according to an embodiment of this application. The data processing method may be performed by a service server (for example, the service server 100 shown in FIG. 1), or may be performed through interaction between the service server and a terminal device (for example, the terminal device 200a shown in FIG. 1). For ease of understanding, an example in which the method is performed by a service server is used for description in this embodiment of this application. In other words, the service server is used as a computer device. As shown in FIG. 6, the data processing method may include at least the following operation S201 to operation S207.

[00110] Operation S201: Obtain a preloading request sent by a terminal device, and obtain, based on the preloading request, a first preloading usage probability corresponding to the terminal device, the first preloading usage probability indicating a probability that a target object corresponding to the terminal device interacts with preloaded content.

[00111] In some embodiments, B historical preloaded content lists generated for the terminal device within a first time period are obtained from a database. B is a positive integer. The first time period is earlier than a second time period, and the first time period is adjacent to the second time period. Historical list scores respectively corresponding to B historical preloaded content lists are

obtained from the database, and the B historical list scores are summed to obtain a first score. C historical preloaded content lists outputted by the terminal device are determined from the B historical preloaded content lists. C is a natural number less than or equal to B. The first preloading usage probability is determined based on the first score and the C historical preloaded content lists.

[00112] A process of determining the first preloading usage probability based on the first score and the C historical preloaded content lists may include: summing historical list scores respectively corresponding to the C historical preloaded content lists, to obtain a second score; and determining a score ratio between the first score and the second score, and determining the first preloading usage probability based on the score ratio.

[00113] As described in operation S101 in the embodiment corresponding to FIG. 3, if the terminal device sends the preloading request to the service server within a second-type time period, that is, a content output peak time period, the service server obtains the first preloading usage probability corresponding to the terminal device. The method provided in this embodiment of this application may be applied to a plurality of projects and product applications, including short videos, long videos, image-text content, and the like. Because different product applications focus on data of different dimensions, the method for determining the first preloading usage probability is not limited in embodiments of this application, and may be set based on an actual application scenario.

[00114] This embodiment of this application describes a feasible method for determining a preloading usage probability. When updating a preloading usage probability corresponding to the terminal device, the service server obtains, from the database, the B historical preloaded content lists generated for the terminal device within the first time period. As can be learned from the descriptions of operation S101 in the foregoing embodiment corresponding to FIG. 3 above, the preloading usage probability corresponding to the terminal device is time-sensitive. In other words, the service server updates the preloading usage probability of the terminal device based on the first update period. The first time period may be understood as a time period corresponding to one first update period. For example, if the first update period is half an hour, the service server updates the preloading usage probability of the terminal device once every half hour. For example, the service server generates the preloading usage probability corresponding to the terminal device at 9:30, and the service server obtains, at 10:00, the B historical preloaded content lists generated for the terminal device within the first time period (from 9:30 to 10:00). In other words, the terminal device sent B preloading requests to the service server within the first time period.

[00115] In some embodiments, the service server obtains, from the database, historical list scores respectively corresponding to the B historical preloaded content lists, and sums the B

historical list scores to obtain the first score. The C historical preloaded content lists outputted by the terminal device are determined from the B historical preloaded content lists. A historical preloaded content list outputted by the terminal device may also be understood as a historical preloaded content list having an output attribute, and is a historical preloaded content list that interacts with a target object. In other words, after a historical preloaded content list is returned to the terminal device, a target object corresponding to the terminal device triggers output of preloaded content in the historical preloaded content list, for example, the target object views the preloaded content in the historical preloaded content list through the terminal device.

[00116] The service server sums the historical list scores respectively corresponding to the C historical preloaded content lists to obtain the second score, and determines the score ratio between the first score and the second score. The service server may determine the score ratio as a current preloading usage probability corresponding to the terminal device, that is, the first preloading usage probability. Alternatively, the first preloading usage probability is determined in combination with data of other dimensions of the application client, for example, the quantity ratio described in operation S101 in the embodiment corresponding to FIG. 3. For example, weighted summation is performed on the quantity ratio and the score ratio to determine the first preloading usage probability, where a sum of weights respectively corresponding to the quantity ratio and the score ratio is equal to 1.

[00117] Operation S202: Generate a target preloaded content list based on a pre-recommended content pool for the terminal device if the first preloading usage probability is less than an online recommendation threshold, a generation timestamp of the pre-recommended content pool being earlier than a generation timestamp of the preloading request.

[00118] Operation S203: Generate an online recommended content set for the terminal device if the first preloading usage probability is greater than or equal to the online recommendation threshold, and generate the target preloaded content list based on the online recommended content set, a generation timestamp of the online recommended content set being later than the generation timestamp of the preloading request.

[00119] In some embodiments, with reference to the descriptions of operation S202 and operation S203, in an application scenario (for example, short video browsing), time consumed by waiting for playback when the terminal device enters a short video feed can be effectively reduced by using the preloaded content in this embodiment of this application. Because content preloaded to the client is not necessarily used by the target object, in other words, the target object may not trigger playback of the content, if the preloaded content is not used, resources consumed by such

preloading process are wasted. Based on this, in this embodiment of this application, real-time recommended resources (that is, the online recommended content) are provided for a target object with a high preloading usage probability, and pre-recommended resources (that is, the pre-recommended content) are provided for a target object with a low preloading usage probability. According to the foregoing resource allocation method, both experience of the target object and recommendation quality can be ensured.

[00120] It can be learned from the foregoing that the pre-recommended content pool is constructed, so that high-quality content can be provided at low costs, and system availability can be effectively ensured. Resource utilization may be maximized based on efficient resource allocation between real-time recommended resources and pre-recommended resources.

[00121] Operation S204: Determine content types respectively corresponding to A pieces of preloaded content in the preloaded content list, A being a positive integer.

[00122] In some embodiments, an application scenario is not limited in embodiments of this application, and a content type is not limited either. Therefore, a content type is not limited, and may be determined based on an actual application scenario. The content type may include but is not limited to the following: a type similar to an output content feature, a friend-published type, a friend-interacted type, and an account-interacted type. Preloaded content that belongs to the type similar to the output content feature is preloaded content whose similarity to the output content feature is greater than the similarity threshold z. Preloaded content that belongs to the friend-published type is preloaded content that is content published by a friend account. Preloaded content that belongs to the friend-interacted type is preloaded content that is content with which a friend account interacts (for example, likes, comments, forwards, or rewards). Preloaded content that belongs to the account-interacted type is preloaded content that is content with which an application account logged in to the application client interacts.

[00123] Operation S205: Determine, for each of the A pieces of preloaded content based on a content type corresponding to the preloaded content, a content score corresponding to the preloaded content.

[00124] In some embodiments, the service server may set a content score for each content type. With reference to the content types shown in operation S204, for example, a content score of the type similar to the output content feature is set to 2, a content score of the friend-published type is set to 3, a content score of the friend-interacted type is set to 1, and a content score of the account-interacted type is set to 2.

[00125] The service server may determine a content score for each piece of preloaded content based on a content type of the preloaded content. For example, the target preloaded content list 20i shown in FIG. 2 includes the video 1, the video 2, and the video 3. The video 1 belongs to the type similar to the output content feature, and therefore a content score for the video 1 is 2. The video 2 belongs to the friend-published type, and therefore a content score for the video 2 is 3. The video 3 belongs to the friend-published type, and therefore a content score for the video 3 is 3.

[00126] Operation S206: Sum content scores respectively corresponding to A pieces of preloaded content, to obtain a list score corresponding to the preloaded content list, the list score being configured for determining a second preloading usage probability corresponding to the second time period.

[00127] In some embodiments, with reference to descriptions of operation S205, the service server determines content scores, that is, 2, 3, and 3, respectively corresponding to the three videos in the target preloaded content list 20i shown in FIG. 2, and sums the three content scores to obtain a list score corresponding to the target preloaded content list 20i, that is, 2+3+3=8. The foregoing method for determining a historical list score corresponding to a historical preloaded content list is the same as the method for determining a list score corresponding to a target preloaded content list.

[00128] Operation S207: Associatively store, in the database, the list score and a device identifier indicating the terminal device.

[00129] In some embodiments, beneficial effects brought by embodiments of this application may be summarized as follows.

[00130] 1. Feed stream (for example, a short video) experience and availability are improved at very low costs, and high-quality recommended content is maintained.

[00131] 2. A method for generating a pre-recommended content pool and application of the pre-recommended content pool are provided, so that massive preloading requests are processed at low costs, thereby ensuring high-quality feed stream experience.

[00132] 3. A resource allocation method based on a preloading usage probability is provided, so that a target object with a high usage probability obtains higher-quality resources, thereby further improving recommendation quality without increasing costs.

[00133] 4. A method for using a pre-recommended content pool as a fallback when real-time recommendation fails is provided, thereby ensuring system availability.

[00134] It can be learned from the above that this embodiment of this application provides a resource allocation method based on the first preloading usage probability. For a terminal device

corresponding to the first preloading usage probability that is greater than or equal to the online recommendation threshold, the online recommended content set generated in real time is provided, and the preloaded content list generated in real time can further be provided. For a terminal device corresponding to the first preloading usage probability that is less than the online recommendation threshold, the preloaded content is allocated from the pre-recommended content pool (that is, a recommended content pool generated in advance). Therefore, resource costs for real-time computation can be reduced by using the pre-recommended content pool. The first preloading usage probability is compared with the online recommendation threshold, so that high-quality preloaded content (that is, recommended content generated in real time) can be allocated to a target object with a high usage probability, thereby improving content recommendation quality and improving utilization of the online recommended content.

[00135] FIG. 7 is a third schematic flowchart of a data processing method according to an embodiment of this application. The method may be performed by a service server (for example, the service server 100 shown in FIG. 1), or may be performed through interaction between the service server and a terminal device (for example, the terminal device 200a shown in FIG. 1). For ease of understanding, an example in which the method is performed by a service server is used for description in this embodiment of this application. In other words, the service server is used as a computer device. As shown in FIG. 7, the method may include at least operation S301 to operation S306.

[00136] Operation S301: Obtain, from a database, D preloading usage probabilities of a full set of terminal devices within a first time period, the full set of terminal devices including a terminal device, and the D preloading usage probabilities including a first preloading usage probability.

[00137] In some embodiments, the full set of terminal devices refer to all terminal devices that perform content consumption within the first time period. The first time period may be understood as a time period corresponding to a first update period. The service server updates, based on the first update period, the preloading usage probabilities respectively corresponding to the full set of terminal devices. For example, if the first update period is half an hour, the service server updates, every half hour, the preloading usage probabilities respectively corresponding to the full set of terminal devices. For example, if a current preloading usage probability of the terminal device is generated at 9:30, the service server updates the preloading usage probability of the terminal device at 10:00. Therefore, the foregoing first time period may be from 9:30 to 10:00.

[00138] A total quantity of the full set of terminal devices is equal to D.

[00139] Operation S302: Sort the D preloading usage probabilities to obtain D sorted preloading usage probabilities.

[00140] In some embodiments, the service server may sort the D preloading usage probabilities in descending order to obtain the D sorted preloading usage probabilities. In other words, a preloading usage probability ranked earlier is greater than or equal to a preloading usage probability ranked later.

[00141] Operation S303: Determine a first online recommendation quota corresponding to system time, and determine an online recommendation threshold from the sorted D preloading usage probabilities based on the first online recommendation quota.

[00142] In some embodiments, a second online recommendation quota corresponding to the system time is determined, and difference calculation is performed on a preset online recommendation total quota and the second online recommendation quota to obtain the first online recommendation quota. E preloading usage probabilities whose quantity is equal to the first online recommendation quota are obtained from the D sorted preloading usage probabilities. E is a positive integer less than or equal to D. A smallest preloading usage probability among the E preloading usage probabilities is determined as the online recommendation threshold.

[00143] In this embodiment of this application, real-time recommended resources (that is, online recommended content) are dynamically allocated, to utilize the real-time recommended resources to a maximum extent. In some embodiments, a total real-time recommendation quota, that is, a fixed online recommendation total quota, is defined. The service server determines the second online recommendation quota based on a second update period. For example, the second update period is one second or one minute. In other words, the service server collects statistics on a total quantity of non-preloading requests obtained in a current second (or current minute), and determines the total quantity of non-preloading requests as the second online recommendation quota in the next second (or next minute). A non-preloading request is a request generated after the terminal device enters a content output page and when a content output request exists. Content returned by the service server for the non-preloading request is referred to as non-preloaded content. The non-preloaded content is online recommended content, and therefore is also referred to as non-preloaded online recommended content in this embodiment of this application. The second online recommendation quota is a total amount of non-preloaded online recommended content, the non-preloaded online recommended content is content pulled by the terminal device after the terminal device enters a content playback page, and first online recommended content is content pulled by the terminal device before the terminal device enters the content playback page.

[00144] The second online recommendation quota and the first online recommendation quota need to satisfy Formula (1). Refer also to FIG. 8. FIG. 8 is an example stacked line diagram of a first online recommended content quota according to an embodiment of this application. A sum of the second online recommendation quota and the first online recommendation quota is less than or equal to the online recommendation total quota. Specifically, within a content output off-peak time period (which is also referred to as a first-type time period in this embodiment of this application), because a sum of quantities respectively corresponding to preloading requests and non-preloading requests is small, the sum of the second online recommendation quota and the first online recommendation quota may be less than the online recommendation total quota. Consequently, utilization of system resources corresponding to the service server is low. In this case, for the preloading request, the service server returns the online recommended content as the preloaded content to the terminal device within the content output off-peak time period. In other words, pre-recommended content is not used (which is represented as preloaded pre-recommendation in FIG. 8).

[00145] Within a content output peak time period (which is also referred to as a second-type time period in this embodiment of this application), there are many online target objects, and a sum of quantities respectively corresponding to the preloading requests and the non-preloading requests is greater than the online recommendation total quota. Therefore, the service server sets a real-time recommendation quantity for preloading requests, and dynamically adjusts the real-time recommendation quantity based on a real-time recommendation quantity for non-preloading requests, to utilize real-time recommended resources to a maximum extent. As shown in FIG. 8, within the content output peak time period, the sum of the first online recommendation quota and the second online recommendation quota is equal to the fixed online recommendation total quota.

[00146] Operation S304: Obtain a preloading request sent by the terminal device, and obtain, based on the preloading request, a first preloading usage probability corresponding to the terminal device, the first preloading usage probability indicating a probability that a target object corresponding to the terminal device interacts with preloaded content.

[00147] Operation S305: Generate a target preloaded content list based on a pre-recommended content pool for the terminal device if the first preloading usage probability is less than an online recommendation threshold, a generation timestamp of the pre-recommended content pool being earlier than a generation timestamp of the preloading request.

[00148] Operation S306: Generate an online recommended content set for the terminal device if the first preloading usage probability is greater than or equal to the online recommendation

threshold, and generate the target preloaded content list based on the online recommended content set, a generation timestamp of the online recommended content set being later than the generation timestamp of the preloading request.

[00149] For a specific implementation process of operation S304 to operation S306, refer to operation S101 to operation S104 in the embodiment corresponding to FIG. 3. Details are not described herein again.

[00150] It can be learned from the above that this embodiment of this application provides a resource allocation method based on the first preloading usage probability. For a terminal device corresponding to the first preloading usage probability that is greater than or equal to the online recommendation threshold, the online recommended content set generated in real time is provided, and the preloaded content list generated in real time can further be provided. For a terminal device corresponding to the first preloading usage probability that is less than the online recommendation threshold, the preloaded content is allocated from the pre-recommended content pool (that is, a recommended content pool generated in advance). Therefore, resource costs for real-time computation can be reduced by using the pre-recommended content pool. The first preloading usage probability is compared with the online recommendation threshold, so that high-quality preloaded content (that is, recommended content generated in real time) can be allocated to a target object with a high usage probability, thereby improving content recommendation quality and improving utilization of the online recommended content. In addition, this embodiment of this application provides a method for jointly and dynamically allocating the online recommendation total quota between the preloading requests and the non-preloading requests, thereby improving utilization of online recommended resources.

[00151] FIG. 9 is a fourth schematic flowchart of a data processing method according to an embodiment of this application. The method may be performed by a service server (for example, the service server 100 shown in FIG. 1), or may be performed through interaction between the service server and a terminal device (for example, the terminal device 200a shown in FIG. 1). For ease of understanding, an example in which the method is performed by a service server is used for description in this embodiment of this application. In other words, the service server is used as a computer device. As shown in FIG. 9, the method includes at least the following operations.

[00152] Operation S401: Obtain a historical online recommended content set generated for a terminal device, and determine, from the historical online recommended content set, first historical online recommended content having an output attribute.

[00153] In some embodiments, the first historical online recommended content having the output attribute is first historical online recommended content outputted by the terminal device. To implement high-quality feed stream experience, various multimedia applications perform preloading before a target object consumes multimedia content. The preloading includes preloading a recommended feed list (which is also referred to as a preloaded content list in this application), pre-downloading a content header, pre-rendering a player, and the like. Content is loaded before the target object views the content, so that when the target object consumes the content, the application client does not experience stuttering or slow loading, thereby improving consumption experience of the target object.

[00154] An embodiment of this application provides a low-cost method for improving feed stream experience and availability while maintaining high recommendation quality. The historical online recommended content set is an online recommended content set generated in response to a preloading request obtained within a historical time period (a historical preloading request for short). A total quantity of historical online recommended content sets is not limited in embodiments of this application, and may be set based on an actual application scenario. An optional implementation is as follows: The historical online recommended content set may include an online recommended content set generated in response to a preloading request obtained within a historical time period, and may further include an online recommended content set generated in response to a non-preloading request obtained within the historical time period.

[00155] For example, if current system time is 12:00 on May 6, 2024, an online recommended content set generated for the terminal device within a week (or one day, or three days) before 12:00 on May 6, 2024 may be referred to as a historical online recommended content set.

[00156] The service server determines, from the historical online recommended content set, the first historical online recommended content having an output attribute. The first historical online recommended content having an output attribute is online recommended content that has been consumed by the target object. For example, the online recommended content is returned to the terminal device as preloaded content (or non-preloaded content), and the target object consumes the online recommended content on the terminal device. Therefore, the online recommended content has the output attribute.

[00157] Operation S402: Add second historical online recommended content to an initial pre-recommended content pool, the second historical online recommended content being recommended content in the historical online recommended content set other than the first historical online recommended content.

[00158] In some embodiments, corresponding to the first historical online recommended content, the second historical online recommended content is online recommended content that has not been consumed by the target object, and therefore has a to-be-output attribute.

[00159] Operation S403: Generate a pre-recommended content pool based on the initial pre-recommended content pool.

[00160] In some embodiments, offline recommendation processing is performed on the terminal device within a first-type time period to obtain offline recommended content. The offline recommended content is added to the initial pre-recommended content pool, and the initial pre-recommended content pool added with which the offline recommended content is determined as the pre-recommended content pool.

[00161] In this application, the pre-recommended content is constructed from two types of content: online computed content triggered by a preloading request, and offline precomputed content. The online computed content includes an online recommended content set generated in response to a preloading request, and may further include an online recommended content set generated in response to a non-preloading request. The offline precomputed content is offline recommended content generated by the service server through offline computation for the terminal device within a content output off-peak time period. The offline recommendation processing does not need to be triggered by the target object, and may be set by the service server.

[00162] It can be learned with reference to operation S401 to operation S403 that, in this embodiment of this application, a feed list generated in a pulling process (which may include a preloading process and a non-preloading process) is stored as the initial pre-recommended content pool, and the initial pre-recommended content pool is supplemented with feeds generated through offline computation within the content output off-peak time period. In this way, sufficient and high-quality personalized pre-recommended content can be obtained at low costs.

[00163] It can be learned from the above that, in embodiments of this application, a preloading mechanism is implemented, a pre-recommended content pool is constructed, and real-time recommended content or pre-recommended content may be allocated based on a probability that a preloading request is used, so that the target object consumes real-time recommended content to a maximum extent. In this way, timeliness of content recommended to the target object and content recommendation quality can be ensured, thereby improving attractiveness of the recommended content to the target object, encouraging the target object to participate in content creation and interaction, improving satisfaction of the target object with a recommendation platform, and improving user stickiness.

[00164] Refer also to FIG. 10. FIG. 10 is an example diagram of classification of loaded content according to an embodiment of this application. Loaded content may be classified into preloaded content and non-preloaded content. The preloaded content is content pulled in advance by the terminal device before the terminal device enters a content consumption page. The non-preloaded content is content pulled in real time by the terminal device when the terminal device is consuming content after entering the content consumption page. Therefore, all the non-preloaded content is online recommended content. The preloaded content is classified into two types. One type is pre-recommended content, and the other type is online recommended content. When to use the pre-recommended content as the preloaded content and when to use the online recommended content as the preloaded content can be found in the descriptions of the embodiment corresponding to FIG. 3 above. The pre-recommended content may also be classified into two types. One type is historical online recommended content, and the other type is offline recommended content. For meanings and generation of the foregoing two types of content, refer to the descriptions of the foregoing operations.

[00165] Operation S404: Obtain a preloading request sent by the terminal device, and obtain, based on the preloading request, a first preloading usage probability corresponding to the terminal device, the first preloading usage probability indicating a probability that a target object corresponding to the terminal device interacts with preloaded content.

[00166] For a specific implementation process of operation S404, refer to operation S101 in the embodiment corresponding to FIG. 3.

[00167] Operation S405: Generate a target preloaded content list based on a pre-recommended content pool for the terminal device if the first preloading usage probability is less than an online recommendation threshold, a generation timestamp of the pre-recommended content pool being earlier than a generation timestamp of the preloading request.

[00168] In some embodiments, a pre-recommended content pool of high-quality feeds is used for preloading in a preceding scenario, so that preloading coverage can be improved and waiting time of the target object during browsing can be reduced. In this embodiment of this application, based on a probability that the target object uses preloaded data (that is, a preloading usage probability), online recommended resources (that is, the online recommended content) are allocated to a target object with a high usage probability, and the pre-recommended content is allocated to the target object with a low usage probability, so that the system resources are utilized to a maximum extent to maximize service indicators. In addition, when a recommendation system fails, the

pre-recommended content pool is used to replace real-time recommended content, so that feed stream availability can be fully ensured while high recommendation quality is ensured.

[00169] Operation S406: Generate an online recommended content set for the terminal device if the first preloading usage probability is greater than or equal to the online recommendation threshold, and generate the target preloaded content list based on the online recommended content set, a generation timestamp of the online recommended content set being later than the generation timestamp of the preloading request.

[00170] In some embodiments, refer to FIG. 11 and FIG. 12. FIG. 11 is a schematic diagram of a usage probability of online recommended content under random allocation according to an embodiment of this application, and FIG. 12 is a schematic diagram of a usage probability of online recommended content under probability-based allocation according to an embodiment of this application. As shown in FIG. 11, for preloading requests, online recommended content is used with a probability of 30%, and pre-recommended content is used with a probability of 70%. For example, after obtaining preloading requests respectively sent by 100 terminal devices, that is, after obtaining 100 preloading requests, a service server randomly selects 30 preloading requests from the 100 preloading requests, returns preloaded content generated from the online recommended content for the 30 preloading requests, and returns preloaded content generated from the pre-recommended content for the 70 preloading requests. Because the online recommended content is randomly allocated to the preloading requests, in FIG. 11, only 14% of the preloaded content generated from the online recommended content is used, that is, consumed, and 86% of the preloaded content generated from the online recommended content is not used.

[00171] Refer to FIG. 12. According to this embodiment of this application, for preloading requests, online recommended content is still used with a probability of 30% and pre-recommended content is still used with a probability of 70%. In addition, preloading usage probabilities respectively corresponding to terminal devices that send the preloading requests are obtained, and the preloading usage probabilities are sorted. For example, after obtaining preloading requests respectively sent by 100 terminal devices, that is, obtaining 100 preloading requests, a service server obtains preloading usage probabilities respectively corresponding to the 100 terminal devices and sorts the 100 preloading usage probabilities in descending order. For preloading usage probabilities ranked in the top 30, the service server responds, to terminal devices respectively corresponding to these preloading usage probabilities, with preloaded content generated from the online recommended content. For preloading usage probabilities ranked in the bottom 70, the

service server responds, to terminal devices respectively corresponding to these preloading usage probabilities, with preloaded content generated from the pre-recommended content.

[00172] Because the online recommended content is allocated based on preloading usage probabilities, in FIG. 12, 25% of the preloaded content generated from the online recommended content is used, that is, consumed, and 75% of the preloaded content generated from the online recommended content is not used. Compared with a case in FIG. 11 in which only 14% of the preloaded content generated from the online recommended content is used, FIG. 12 shows an increase of 11%. In other words, in this embodiment of this application, computing resources, storage resources, and network bandwidth can be utilized more properly, thereby avoiding resource waste and improving resource utilization and content usage.

[00173] It can be learned from the above that this embodiment of this application provides a resource allocation method based on the first preloading usage probability. For a terminal device corresponding to the first preloading usage probability that is greater than or equal to the online recommendation threshold, the online recommended content set generated in real time is provided, and the preloaded content list generated in real time can further be provided. For a terminal device corresponding to the first preloading usage probability that is less than the online recommendation threshold, the preloaded content is allocated from the pre-recommended content pool (that is, a recommended content pool generated in advance). Therefore, resource costs for real-time computation can be reduced by using the pre-recommended content pool. The first preloading usage probability is compared with the online recommendation threshold, so that high-quality preloaded content (that is, recommended content generated in real time) can be allocated to a target object with a high usage probability, thereby improving content recommendation quality and improving utilization of the online recommended content.

[00174] FIG. 13 is a schematic diagram of a structure of a data processing apparatus according to an embodiment of this application. The data processing apparatus 1 may be configured to perform the corresponding operations in the method according to embodiments of this application. As shown in FIG. 13, the data processing apparatus 1 may include an obtaining module 11 and a generation module 12.

[00175] The obtaining module 11 is configured to obtain a preloading request sent by a terminal device, and obtain, based on the preloading request, a first preloading usage probability corresponding to the terminal device, the first preloading usage probability indicating a probability that a target object corresponding to the terminal device interacts with preloaded content. The generation module 12 is configured to: generate an online recommended content set for the terminal

device if the first preloading usage probability is greater than or equal to an online recommendation threshold; and generate a target preloaded content list based on the online recommended content set, a generation timestamp of the online recommended content set being later than a generation timestamp of the preloading request.

[00176] In some embodiments, the generation module 12 is further configured to generate the target preloaded content list based on a pre-recommended content pool for the terminal device if the first preloading usage probability is less than the online recommendation threshold, a generation timestamp of the pre-recommended content pool being earlier than the generation timestamp of the preloading request.

[00177] In some embodiments, the obtaining module 11 is further configured to: obtain, at a first moment, the preloading request sent by the terminal device; determine a quantity of application startup times and a quantity of preloading usage times of the terminal device within a first time period, the latest timestamp of the first time period is earlier than the first moment; and determine a quantity ratio between the quantity of application startup times and the quantity of preloading usage times, and determine the first preloading usage probability based on the quantity ratio.

[00178] In some embodiments, the data processing apparatus 1 further includes an associative storage module. The associative storage module is configured to: associatively store, in a database, the first preloading usage probability and a device identifier indicating the terminal device.

[00179] In some embodiments, the obtaining module 11 is further configured to perform the following operations: determining content types respectively corresponding to A pieces of preloaded content in a preloaded content list, A being a positive integer; determining, for each of the A pieces of preloaded content based on a content type corresponding to the preloaded content, a content score corresponding to the preloaded content; summing content scores respectively corresponding to A pieces of preloaded content, to obtain a list score corresponding to the preloaded content list, the list score being configured for determining a second preloading usage probability corresponding to a second time period; and associatively storing, in the database, the list score and the device identifier indicating the terminal device.

[00180] In some embodiments, the obtaining module 11 is further configured to perform the following operations: obtaining, from the database, B historical preloaded content lists generated for the terminal device within a first time period, B being a positive integer, the first time period being earlier than the second time period, and the first time period being adjacent to the second time period; obtain, from the database, historical list scores respectively corresponding to B historical

preloaded content lists, and sum the B historical list scores to obtain a first score; determine, from the B historical preloaded content lists, C historical preloaded content lists outputted by the terminal device, C being a natural number less than or equal to B; and determine the first preloading usage probability based on the first score and the C historical preloaded content lists.

[00181] In some embodiments, the obtaining module 11 determines the first preloading usage probability based on the first score and the C historical preloaded content lists, and is configured to perform the following operations: summing historical list scores respectively corresponding to the C historical preloaded content lists, to obtain a second score; and determining a score ratio between the first score and the second score, and determining the first preloading usage probability based on the score ratio.

[00182] In some embodiments, the obtaining module 11 is further configured to perform the following operations: obtaining, from the database, D preloading usage probabilities of a full set of terminal devices within the first time period, the full set of terminal devices including the terminal device, and the D preloading usage probabilities including the first preloading usage probability; sorting the D preloading usage probabilities to obtain D sorted preloading usage probabilities; and determining a first online recommendation quota corresponding to system time, and determining the online recommendation threshold from the D sorted preloading usage probabilities based on the first online recommendation quota, the first online recommendation quota being a first total amount of content pulled by the terminal device before the terminal device enters a content playback page.

[00183] In some embodiments, the obtaining module 11 determines the first online recommendation quota corresponding to the system time, and is configured to perform the following operations: determining a second online recommendation quota corresponding to the system time, and performing difference calculation on a preset online recommendation total quota and the second online recommendation quota to obtain the first online recommendation quota. The obtaining module determines the online recommendation threshold from the D sorted preloading usage probabilities based on the first online recommendation quota, and is configured to perform the following operation: obtaining, from the D sorted preloading usage probabilities, E preloading usage probabilities whose quantity is equal to the first online recommendation quota, E being a positive integer less than or equal to D; and determining a smallest preloading usage probability among the E preloading usage probabilities as the online recommendation threshold.

[00184] In some embodiments, the obtaining module 11 is further configured to perform the following operations: obtaining a historical online recommended content set generated for the terminal device, and determining, from the historical online recommended content set, first

historical online recommended content outputted by the terminal device; adding second historical online recommended content to an initial pre-recommended content pool, the second historical online recommended content being recommended content in the historical online recommended content set other than the first historical online recommended content; and generating a pre-recommended content pool based on the initial pre-recommended content pool.

[00185] In some embodiments, the generation module 12 generates the pre-recommended content pool based on the initial pre-recommended content pool, and is configured to perform the following operations: performing offline recommendation processing on the terminal device within a first-type time period to obtain offline recommended content; and adding the offline recommended content to the initial pre-recommended content pool, and determining an initial pre-recommended content pool to which the offline recommended content is added as the pre-recommended content pool.

[00186] In some embodiments, the generation module 12 generates the target preloaded content list based on the pre-recommended content pool for the terminal device, and is configured to perform the following operations: determining, from the pre-recommended content pool for the terminal device, F pieces of pre-recommended content that are in a valid state, F being a positive integer; determining recommendation scores respectively corresponding to the F pieces of pre-recommended content, and obtaining G recommendation scores from the F recommendation scores, the G recommendation scores being greater than or equal to H recommendation scores, the H recommendation scores including a recommendation score in the F recommendation scores other than the G recommendation scores, G being a positive integer less than or equal to F, and H being a positive integer less than F; and obtaining, from the F pieces of pre-recommended content, G pieces of pre-recommended content respectively corresponding to the G recommendation scores, and generating the target preloaded content list based on the G pieces of pre-recommended content.

[00187] In some embodiments, the generation module 12 generates the target preloaded content list based on the G pieces of pre-recommended content, and is configured to perform the following operations: generating a first preloaded content list based on the G pieces of pre-recommended content, the G pieces of pre-recommended content including first pre-recommended content and second pre-recommended content; determining a content similarity between the first pre-recommended content and the second pre-recommended content; deleting the first pre-recommended content from the first preloaded content list if the content similarity is greater than or equal to a similarity threshold, to obtain a second preloaded content list; obtaining a largest recommendation score from the H recommendation scores, and obtaining, from the F pieces

of pre-recommended content, third pre-recommended content corresponding to the largest recommendation score; adding the third pre-recommended content to the second preloaded content list, to obtain a third preloaded content list; and generating the target preloaded content list based on the third preloaded content list, a content similarity between any two pieces of preloaded content in the target preloaded content list being less than the similarity threshold.

[00188] In some embodiments, the obtaining module 11 obtains, based on the preloading request, the first preloading usage probability corresponding to the terminal device, and is configured to perform the following operations: determining an obtaining timestamp of the preloading request, generating an online recommended content set for the terminal device if the obtaining timestamp is within the first-type time period, and generating the target preloaded content list based on the online recommended content set; or obtaining the first preloading usage probability corresponding to the terminal device if the obtaining timestamp is within the second-type time period, a total amount of output content corresponding to the second-type time period being greater than a total amount of output content corresponding to the first-type time period.

[00189] In some embodiments, the generation module 12 generates the online recommended content set for the terminal device, and is configured to perform the following operations: determining a total quantity of preloading usage probabilities that are greater than or equal to the online recommendation threshold; generating the online recommended content set for the terminal device based on an output content feature of the terminal device and a candidate content pool corresponding to system time if the total quantity is less than or equal to the first online recommendation quota; or generating the target preloaded content list based on the pre-recommended content pool if the total quantity is greater than the first online recommendation quota.

[00190] It can be learned from the above that this embodiment of this application provides a resource allocation method based on the first preloading usage probability. For a terminal device corresponding to the first preloading usage probability that is greater than or equal to the online recommendation threshold, the online recommended content set generated in real time is provided, and the preloaded content list generated in real time can further be provided. For a terminal device corresponding to the first preloading usage probability that is less than the online recommendation threshold, the preloaded content is allocated from the pre-recommended content pool (that is, a recommended content pool generated in advance). Therefore, resource costs for real-time computation can be reduced by using the pre-recommended content pool. The first preloading usage probability is compared with the online recommendation threshold, so that high-quality preloaded

content (that is, recommended content generated in real time) can be allocated to a target object with a high usage probability, thereby improving content recommendation quality and improving utilization of the online recommended content.

[00191] FIG. 14 is a schematic diagram of a structure of a computer device according to an embodiment of this application. As shown in FIG. 14, the computer device 1000 may include at least one processor 1001, for example, an CPU, at least one network interface 1004, a user interface 1003, a memory 1005, and at least one communication bus 1002. The communication bus 1002 is configured to implement connection communication between these components. In some embodiments, the user interface 1003 may include a display and a keyboard, and in some embodiments, the network interface 1004 may include a standard wired interface and a standard wireless interface (for example, a Wi-Fi interface). The memory 1005 may be a high-speed random access memory (RAM) or a non-volatile memory, for example, at least one magnetic disk memory. In some embodiments, the memory 1005 may alternatively be at least one storage apparatus located away from the processor 1001. As shown in FIG. 14, the memory 1005 used as a computer-readable storage medium may include an operating system, a network communication module, a user interface module, and a device control application program.

[00192] In the computer device 1000 shown in FIG. 14, the network interface 1004 may provide a network communication function. The user interface 1003 is mainly configured to provide an input interface for a user. The processor 1001 may be configured to invoke the device control application program stored in the memory 1005 to implement the following operations:

[00193] obtaining a preloading request sent by a terminal device, and obtaining, based on the preloading request, a first preloading usage probability corresponding to the terminal device, the first preloading usage probability indicating a probability that a target object corresponding to the terminal device interacts with preloaded content;

[00194] generating an online recommended content set for the terminal device if the first preloading usage probability is greater than or equal to an online recommendation threshold, and generating a target preloaded content list based on the online recommended content set, a generation timestamp of the online recommended content set being later than a generation timestamp of the preloading request.

[00195] The computer device 1000 described in this embodiment of this application may perform the descriptions of the data processing method or apparatus in the foregoing embodiments. Details are not described herein again. In addition, descriptions of beneficial effects achieved by using the same method are not described herein again.

[00196] An embodiment of this application further provides a computer-readable storage medium. The computer-readable storage medium stores a computer program. When the computer program is executed by a processor, the descriptions of the data processing method or apparatus in the foregoing embodiments are implemented. Details are not described herein again. In addition, descriptions of beneficial effects achieved by using the same method are not described herein again.

[00197] The foregoing computer-readable storage medium may be the data processing apparatus provided in any one of the foregoing embodiments or an internal storage unit of the foregoing computer device, for example, a hard disk drive or a memory of the computer device. The computer-readable storage medium may alternatively be an external storage device of the computer device, for example, a plug-in hard disk drive, a smart media card (SMC), a secure digital (SD) card, or a flash card that is configured on the computer device. The computer-readable storage medium may alternatively include both the internal storage unit of the computer device and the external storage device. The computer-readable storage medium is configured to store the computer program and other programs and data that are needed by the computer device. The computer-readable storage medium may be further configured to temporarily store data that has been or is to be outputted.

[00198] An embodiment of this application further provides a computer program product. The computer program product includes a computer program, and the computer program is stored on a computer-readable storage medium. A processor of a computer device reads the computer program from the computer-readable storage medium, and the processor executes the computer program, so that the computer device can perform the descriptions of the data processing method or apparatus in the foregoing embodiments. Details are not described herein again. In addition, descriptions of beneficial effects achieved by using the same method are not described herein again.

[00199] The terms "first", "second", and the like in the specification, claims, and accompanying drawings of embodiments of this application are configured for distinguishing different objects, and are not used for describing a specific order. In addition, the term "including", or any other variants thereof is intended to cover a non-exclusive inclusion. For example, a process, a method, an apparatus, a product, or a device including a series of steps or units is not limited to the listed steps or modules, but instead, in some embodiments, includes steps or modules that are not listed, or in some embodiments, includes other steps or units inherent to the process, method, apparatus, product, or device.

[00200] A person of ordinary skill in the art may be aware that, in combination with the examples described in embodiments disclosed in this specification, units and algorithm steps may

be implemented by electronic hardware or a combination of computer software and electronic hardware. To clearly describe interchangeability between the hardware and the software, compositions and operations of each example are generally described in the foregoing descriptions based on functions. Whether the functions are performed by hardware or software depends on particular applications and design constraint conditions of the technical solutions. A person skilled in the art may use different methods to implement the described functions for each particular application, but it shall not be considered that the implementation goes beyond the scope of this application.

[00201] The foregoing disclosure is merely exemplary embodiments of this application, and it is clear that, the scope of the claims of this application is not limited thereto. Therefore, any equivalent modification made according to the claims of this application shall fall within the scope of this application.

CLAIMS

1. A data processing method, applied to a computer device, and comprising:

obtaining a preloading request sent by a terminal device, and determining, based on the preloading request, a first preloading usage probability corresponding to the terminal device, the first preloading usage probability indicating a probability that a target object corresponding to the terminal device interacts with preloaded content;

generating an online recommended content set for the terminal device if the first preloading usage probability is greater than or equal to an online recommendation threshold; and

generating a target preloaded content list based on the online recommended content set, a generation timestamp of the online recommended content set being later than a generation timestamp of the preloading request.

2. The method according to claim 1, further comprising:

generating the target preloaded content list based on a pre-recommended content pool for the terminal device if the first preloading usage probability is less than the online recommendation threshold, a generation timestamp of the pre-recommended content pool being earlier than the generation timestamp of the preloading request.

3. The method according to claim 1 or 2, wherein the obtaining a preloading request sent by a terminal device, and determining, based on the preloading request, a first preloading usage probability corresponding to the terminal device comprises:

obtaining, at a first moment, the preloading request sent by the terminal device;

determining a quantity of application startup times and a quantity of preloading usage times of the terminal device within a first time period, a latest timestamp of the first time period is earlier than the first moment; and

determining a quantity ratio between the quantity of application startup times and the quantity of preloading usage times, and determining the first preloading usage probability based on the quantity ratio; and

the method further comprises:

associatively storing, in a database, the first preloading usage probability and a device identifier indicating the terminal device.

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

determining content types respectively corresponding to A pieces of preloaded content in the preloaded content list, A being a positive integer;

for each of the A pieces of preloaded content, determining, based on a content type corresponding to the preloaded content, a content score corresponding to the preloaded content;

summing content scores respectively corresponding to the A pieces of preloaded content, to obtain a list score corresponding to the preloaded content list, the list score being configured for determining a second preloading usage probability corresponding to a second time period; and

associatively storing, in the database, the list score and a device identifier indicating the terminal device.

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

obtaining, from a database, B historical preloaded content lists generated for the terminal device within a first time period, B being a positive integer, the first time period being earlier than a second time period, and the first time period being adjacent to the second time period;

obtaining, from the database, historical list scores respectively corresponding to the B historical preloaded content lists, and summing the B historical list scores to obtain a first score;

determining, from the B historical preloaded content lists, C historical preloaded content lists outputted by the terminal device, C being a natural number less than or equal to B; and

determining the first preloading usage probability based on the first score and the C historical preloaded content lists.

6. The method according to any one of claims 1 to 5, wherein the determining the first preloading usage probability based on the first score and the C historical preloaded content lists comprises:

summing historical list scores respectively corresponding to the C historical preloaded content lists, to obtain a second score; and

determining a score ratio between the first score and the second score, and determining the first preloading usage probability based on the score ratio.

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

obtaining, from a database, D preloading usage probabilities of a full set of terminal devices within a first time period, the full set of the terminal devices comprising the terminal device, and the D preloading usage probabilities comprising the first preloading usage probability;

sorting the D preloading usage probabilities to obtain D sorted preloading usage probabilities; and

determining a first online recommendation quota corresponding to system time, and determining the online recommendation threshold from the D sorted preloading usage probabilities based on the first online recommendation quota, the first online recommendation quota being a first total amount of content pulled by the terminal device before the terminal device enters a content playback page.

8. The method according to any one of claims 1 to 7, wherein the determining a first online recommendation quota corresponding to system time comprises:

determining a second online recommendation quota corresponding to the system time, and performing difference calculation on a preset online recommendation total quota and the second online recommendation quota to obtain the first online recommendation quota, the second online recommendation quota being a second total amount of content pulled by the terminal device after the terminal device enters the content playback page; and

the determining the online recommendation threshold from the D sorted preloading usage probabilities based on the first online recommendation quota comprises:

obtaining, from the D sorted preloading usage probabilities, E preloading usage probabilities whose quantity is equal to the first online recommendation quota, E being a positive integer less than or equal to D; and

determining a smallest preloading usage probability among the E preloading usage probabilities as the online recommendation threshold.

9. The method according to any one of claims 1 to 8, further comprising:

obtaining a historical online recommended content set generated for the terminal device, and determining, from the historical online recommended content set, first historical online recommended content outputted by the terminal device;

adding second historical online recommended content to an initial pre-recommended content pool, the second historical online recommended content being recommended content in the historical online recommended content set other than the first historical online recommended content; and

generating a pre-recommended content pool based on the initial pre-recommended content pool.

10. The method according to any one of claims 1 to 9, wherein the generating a pre-recommended content pool based on the initial pre-recommended content pool comprises:

performing offline recommendation processing on the terminal device within a first-type time period to obtain offline recommended content; and

adding the offline recommended content to the initial pre-recommended content pool, and determining, as the pre-recommended content pool, the initial pre-recommended content pool added with the offline recommended content.

11. The method according to any one of claims 1 to 10, wherein the generating the target preloaded content list based on a pre-recommended content pool for the terminal device comprises:

determining, from the pre-recommended content pool for the terminal device, F pieces of pre-recommended content that are in a valid state, F being a positive integer;

determining recommendation scores respectively corresponding to the F pieces of pre-recommended content, and obtaining G recommendation scores from the F recommendation scores, the G recommendation scores being greater than or equal to H recommendation scores, the H recommendation scores comprising a recommendation score in the F recommendation scores other than the G recommendation scores, G being a positive integer less than or equal to F, and H being a positive integer less than F; and

obtaining, from the F pieces of pre-recommended content, pre-recommended content respectively corresponding to the G recommendation scores, and generating the target preloaded content list based on the G pieces of pre-recommended content.

12. The method according to any one of claims 1 to 11, wherein the generating the target preloaded content list based on the G pieces of pre-recommended content comprises:

generating a first preloaded content list based on the G pieces of pre-recommended content, the G pieces of pre-recommended content comprising first pre-recommended content and second pre-recommended content;

determining a content similarity between the first pre-recommended content and the second pre-recommended content;

deleting the first pre-recommended content from the first preloaded content list if the content similarity is greater than or equal to a similarity threshold, to obtain a second preloaded content list;

obtaining a largest recommendation score from the H recommendation scores, and obtaining, from the F pieces of pre-recommended content, third pre-recommended content corresponding to

the largest recommendation score;

adding the third pre-recommended content to the second preloaded content list, to obtain a third preloaded content list; and

generating the target preloaded content list based on the third preloaded content list, a content similarity between any two pieces of preloaded content in the target preloaded content list being less than the similarity threshold.

13. The method according to any one of claims 1 to 12, wherein the obtaining, based on the preloading request, a first preloading usage probability corresponding to the terminal device comprises:

determining an obtaining timestamp of the preloading request, generating the online recommended content set for the terminal device if the obtaining timestamp is within a first-type time period, and generating the target preloaded content list based on the online recommended content set; or

obtaining the first preloading usage probability corresponding to the terminal device if the obtaining timestamp is within a second-type time period, a total amount of output content corresponding to the second-type time period being greater than a total amount of output content corresponding to the first-type time period.

14. The method according to any one of claims 1 to 13, wherein the generating an online recommended content set for the terminal device comprises:

determining a total quantity of preloading usage probabilities that are greater than or equal to the online recommendation threshold; and

generating the online recommended content set for the terminal device based on an output content feature of the terminal device and a candidate content pool corresponding to system time if the total quantity is less than or equal to a first online recommendation quota; or

generating the target preloaded content list based on a pre-recommended content pool if the total quantity is greater than the first online recommendation quota.

15. A data processing apparatus, comprising:

an obtaining module, configured to: obtain a preloading request sent by a terminal device, and obtain, based on the preloading request, a first preloading usage probability corresponding to the terminal device, the first preloading usage probability indicating a probability that a target object corresponding to the terminal device interacts with preloaded content; and

a generation module, configured to: generate an online recommended content set for the terminal device if the first preloading usage probability is greater than or equal to an online recommendation threshold; and generate a target preloaded content list based on the online recommended content set, a generation timestamp of the online recommended content set being later than a generation timestamp of the preloading request.

16. A computer device, comprising a processor, a memory, and a network interface,

the processor being connected to the memory and the network interface, the network interface being configured to provide a data communication function, the memory being configured to store a computer program, and the processor being configured to invoke the computer program to cause the computer device to perform the method according to any one of claims 1 to 14.

17. A computer-readable storage medium, having a computer program stored thereon, the computer program being loaded and executed by a processor, to cause a computer device having the processor to perform the method according to any one of claims 1 to 14.

18. A computer program product, comprising a computer program, the computer program being stored in a computer-readable storage medium, and the computer program being read and executed by a processor to cause a computer device having the processor to perform the method according to any one of claims 1 to 14.