

METHOD AND APPARATUS FOR PREDICTING NODE STATE

[0001] This application claims priority to Chinese Patent Application No. 202010409392.2, filed on May 14, 2020 and entitled "METHOD AND APPARATUS FOR PREDICTING NODE STATE", which is incorporated herein by reference in its entirety.

5

TECHNICAL FIELD

[0002] This application relates to the field of artificial intelligence technologies, and specifically, to a method and an apparatus for predicting a node state.

BACKGROUND

[0003] In some networks, a current state of a node is related not only to time, but also to space in which the node is located. For example, in a cellular network, traffic carried by a base station in daytime is different from traffic carried by the base station at night, that is, traffic carried by a base station is correlated with time; and in addition, traffic carried by a base station located at a center of the network is different from traffic carried by a base station located at an edge of the network, that is, traffic carried by a base station is correlated with space. To facilitate network management and maintenance, a state in a future period of time needs to be predicted for a node in a network.

[0004] A method for predicting a node state is to obtain a preset network topology view and predict a state of a node in a network based on the preset network topology view and some network-related time information. This method is applicable to a network whose topology structure stays in a fixed state for a long time. For a network whose topology structure changes rapidly with time, such as a cellular network, an effect achieved by applying the above method is less satisfactory.

SUMMARY

[0005] This application provides a method and an apparatus for predicting a node state, so as to increase accuracy of node state prediction in a network with a dynamic topological relationship.

[0006] According to a first aspect, a method for predicting a node state is provided, including: obtaining static graphs and dynamic graphs of a plurality of nodes in a target network, where the static graphs and the dynamic graphs are all topology views; generating spatial feature data of the plurality of nodes based on the static graphs and the dynamic graphs; obtaining time feature data of the plurality of nodes; and obtaining a predicted state of a target node in a target time range based on the spatial feature data and the time feature data, where the target node is any node in the plurality of nodes.

[0007] In a process of node state prediction, a dynamic spatial feature is used in addition to a time feature and a static spatial feature. This increases accuracy of node state prediction in a network with a dynamic topological relationship, so that the network is more intelligent in node state prediction.

[0008] Optionally, the obtaining time feature data of the plurality of nodes includes: obtaining observation data of the plurality of nodes in a first historical time range, where the first historical time range is a time range before the target time range; slicing the observation data in a time dimension to obtain time slice data, where a data amount of the time slice data is less than a data amount of the observation data; and obtaining the time feature data based on the time slice data.

[0009] The time slice data is partial data selected from a complete observation data set. Compared with the complete observation data set, the time slice data has a smaller data amount, and performing prediction by using the time slice data can reduce a time required for prediction.

[0010] Optionally, the time slice data includes: observation data in a time range that is in the first historical time range and that is adjacent to the target time range.

[0011] Optionally, the time slice data includes: observation data in a specific time range in the first historical time range, where the specific time range is located in a first period, the target time range is located in a second period, and the specific time range includes a time range that corresponds to the target time range and that is in the first period.

[0012] The adjacent time range is close to the target time range, and the specific time range and the target time range are located at corresponding time-domain locations in different periods. Therefore, the observation data in the adjacent time range and that in the specific time range have a relatively strong correlation with a node state in the target time range. Performing prediction by using the observation data in the adjacent time range and/or that in the specific time range is beneficial to improving prediction accuracy.

[0013] Optionally, the obtaining static graphs and dynamic graphs of a plurality of nodes includes: obtaining observation data of the plurality of nodes in a second historical time range, where the second historical time range is a time range before the target time range; and constructing the static graphs and the dynamic graphs based on the observation data.

5 **[0014]** The second historical time range may be the same as or different from the first historical time range. A latest spatial feature can be obtained in time by constructing the static graphs and the dynamic graphs by using the observation data.

[0015] Optionally, the observation data includes at least one of the following data: meteorological data, network topology data, traffic data, voice data, signaling data, point of interest
10 POI data, major-event data, and holiday data.

[0016] The network topology data, the traffic data, the voice data, the signaling data, and the POI data may be used to construct the static graphs and the dynamic graphs, to help obtain the spatial feature data. The meteorological data, the major-event data, and the holiday data may be used to obtain the time feature data. The foregoing observation data used to obtain the spatial
15 feature data may also be used to obtain the time feature data. The foregoing data has different sources and different data structures, and may be referred to as multi-source heterogeneous data. The time feature data and the spatial feature data are obtained from the multi-source heterogeneous data, and this enables a temporal model and a spatial model to learn richer knowledge, thereby increasing prediction accuracy.

20 **[0017]** Optionally, the method further includes: obtaining a true state of the target node in the target time range; and training a temporal model and a spatial model based on the true state and the predicted state, where the temporal model is used to output the time feature data based on input observation data, and the spatial model is used to output the spatial feature data based on the static graphs and the dynamic graphs that are input.

25 **[0018]** Through training, prediction accuracy of the spatial model and the temporal model can be increased.

[0019] Optionally, the static graphs belong to first-type topology views, the dynamic graphs belong to second-type topology views, and a topological relationship change rate of the first-type topology views is less than a topological relationship change rate of the second-type topology
30 views.

[0020] Optionally, the target network is a communications network, the static graphs are

topology views representing a traffic-based topological relationship of the plurality of nodes, and the dynamic graphs are topology views representing a physical-line-based topological relationship of the plurality of nodes.

[0021] According to a second aspect, this application provides an apparatus for predicting a node state, including units configured to perform the method according to the first aspect. The apparatus may be a terminal device or a server, or may be a chip in a terminal device or a server. The apparatus may include an input unit and a processing unit.

[0022] When the apparatus is a terminal device or a server, the processing unit may be a processor, and the input unit may be a transceiver. The terminal device may further include a storage unit, and the storage unit may be a memory. The storage unit is configured to store an instruction, and the processing unit executes the instruction stored in the storage unit, so that the terminal device performs the method according to the first aspect.

[0023] When the apparatus is a chip in a terminal device or a server, the processing unit may be a processing unit inside the chip, and the input unit may be an input/output interface, a pin, a circuit, or the like. The processing unit executes an instruction stored in a storage unit, so that the chip performs the method according to the first aspect. The storage unit may be a storage unit (for example, a register or a buffer) in the chip, or may be a storage unit (for example, a read-only memory or a random access memory) located outside the chip.

[0024] According to a third aspect, this application 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 processor is enabled to perform the method according to the first aspect.

[0025] According to a fourth aspect, this application provides a computer program product. The computer program product includes computer program code. When the computer program code is run by a processor, the processor is enabled to perform the method according to the first aspect.

BRIEF DESCRIPTION OF DRAWINGS

[0026] FIG. 1 shows a network applicable to this application;

[0027] FIG. 2 is a schematic diagram of dynamic graphs and static graphs according to this

application;

[0028] FIG. 3 is a schematic diagram of a method for predicting a node state according to this application;

[0029] FIG. 4 is a schematic diagram of historical traffic according to this application;

5 [0030] FIG. 5 is a schematic diagram of a method for obtaining time slice data according to this application;

[0031] FIG. 6 is a schematic diagram of a method for predicting a base station state according to this application;

[0032] FIG. 7 is a schematic diagram of an apparatus for predicting a node state according to
10 this application; and

[0033] FIG. 8 is a schematic diagram of a device for predicting a node state according to this application.

DESCRIPTION OF EMBODIMENTS

[0034] The following describes technical solutions of this application with reference to
15 accompanying drawings.

[0035] FIG. 1 shows a network applicable to this application. A network 100 includes a plurality of base stations. A base station may be referred to as a node. There are two types of connection relationships between the base stations: one is a physical line and the other is traffic. The traffic between the base stations and the base stations jointly form one type of topology view
20 of the network 100, and the physical line between the base stations and the base stations jointly form the other type of topology view of the network 100. Affected by a use behavior and a movement behavior of a user, traffic carried by each base station has a relatively strong correlation with a time dimension and a space dimension. Therefore, a traffic-based topology view continuously changes with time. Such a topology view with a topological relationship changing
25 quickly with time may be referred to as a dynamic graph. The physical line is not affected by the use behavior and the movement behavior of the user. A topological relationship of such a topology view changes slowly with time, and therefore, the topology view may be referred to as a static graph.

[0036] FIG. 2 shows dynamic graphs and static graphs of the network 100. Black dots denote

base stations. In the dynamic graphs, lines between the black dots denote traffic. In the static graphs, lines between the black dots denote physical lines.

[0037] In addition to the network 100, other networks that have dynamic graphs and static graphs are all applicable to this application.

5 **[0038]** For example, in a power grid, substations may be considered as nodes. An electricity transmission line between the substations and the substations jointly form a static graph of the power grid, and a current between the substations and the substations jointly form a dynamic graph of the power grid.

[0039] For another example, in a social network, users may be considered as nodes. A
10 relationship between friendly users and these users jointly form a static graph of the social network, and a relationship between stranger users and these users jointly form a dynamic graph of the social network.

[0040] The following describes the technical solutions of this application in detail by using the network 100 as an example.

15 **[0041]** As shown in FIG. 3, a method 300 includes the following steps.

[0042] S310: Obtain static graphs and dynamic graphs of a plurality of nodes in a target network, where the static graphs and the dynamic graphs are all topology views.

[0043] The static graphs and the dynamic graphs are undirected unweighted graphs. $G=(V,E)$ may be used to denote an undirected unweighted graph, where V denotes a
20 collection of nodes, and E denotes a collection of connection lines between the nodes. A connection matrix $A(i,j)$ may be used to denote a connection relationship between a node i and a node j . When $A(i,j)=1$, it indicates that a connection relationship exists between the node i and the node j ; or when $A(i,j)=0$, it indicates that no connection relationship exists between the node i and the node j .

25 **[0044]** An apparatus for performing the method 300 may construct connection matrices corresponding to different graphs by using a plurality of methods.

[0045] For example, when longitudes and latitudes of the base stations are known, distances between the base stations may be determined, and then connection matrices between the base stations may be determined based on formula (1), so as to further determine the static graphs.

$$A_{SD}(i, j) = F \left(e^{\left(-\frac{d_{ij}^2}{\sigma^2} \right)} \right) \quad (1)$$

[0046] $A_{SD}(i, j)$ denotes a connection relationship between a base station i and a base station j that is determined based on a distance. When $A_{SD}(i, j) = 1$, it indicates that a connection relationship exists the base station i and the base station j , or when $A_{SD}(i, j) = 0$, it indicates that

5 no connection relationship exists between the base station i and the base station j . d_{ij} denotes a distance between the base station i and the base station j . e denotes a natural constant. σ is a hyperparameter, which is used to control a value of a distance between base stations. F is a

mapping function, which is used to map $e^{\left(-\frac{d_{ij}^2}{\sigma^2} \right)}$ to 1 or 0.

[0047] For another example, when historical traffic of the base stations is known, a similarity
10 may be calculated by using the historical traffic, and then the connection matrices between the base stations may be determined based on formula (2), so as to further determine the static graphs.

$$A_{FS}(i, j) = F \left(\rho_{x_w(i), x_w(j)} \right) \quad (2)$$

[0048] $A_{FS}(i, j)$ denotes a connection relationship between a base station i and a base station j that is determined based on the historical traffic. When $A_{FS}(i, j) = 1$, it indicates that a connection

15 relationship exists the base station i and the base station j , or when $A_{FS}(i, j) = 0$, it indicates that

no connection relationship exists between the base station i and the base station j . $x_w(i)$ denotes weekly average traffic data of the base station i . $x_w(j)$ denotes weekly average traffic data of the base station j . $\rho_{x_w(i), x_w(j)}$ denotes a Pearson (Pearson) correlation coefficient between $x_w(i)$ and $x_w(j)$. F is a mapping function, which is used to map $\rho_{x_w(i), x_w(j)}$ to 1 or 0.

20 **[0049]** The method for determining a static graph is described above, and a method for determining a dynamic graph is described below.

[0050] For example, a static graph may be used as an initial graph and then may be updated by using real-time data to obtain a dynamic graph.

[0051] For another example, a dynamic graph may be directly constructed by using real-time

data, as shown in formula (3).

$$A'_{TC}(i, j) = F(\rho_{x_t(i), x_t(j)}) \quad (3)$$

[0052] $A'_{TC}(i, j)$ denotes a connection relationship at moment t between a base station i and a base station j that is determined based on traffic. $x_t(i) = (x_{t-H+1}(i), \dots, x_{t-1}(i))$ denotes a collection of traffic of the base station i at H moments closest to the moment t . $x_t(j) = (x_{t-H+1}(j), \dots, x_{t-1}(j))$ denotes a collection of traffic of the base station j at the H moments closest to the moment t . $\rho_{x_t(i), x_t(j)}$ denotes a Pearson correlation coefficient between $x_t(i)$ and $x_t(j)$. F is a mapping function, which is used to map $\rho_{x_t(i), x_t(j)}$ to 1 or 0.

[0053] Optionally, in the foregoing several examples, a method for mapping F to a variable is shown in formula (4).

$$F(z) = \begin{cases} 1, & |z| \geq \delta \\ 0, & |z| < \delta \end{cases} \quad (4)$$

[0054] z denotes $e^{\left(-\frac{d_{ij}^2}{\sigma^2}\right)}$, $\rho_{x_w(i), x_w(j)}$, or $\rho_{x_t(i), x_t(j)}$. δ is a hyperparameter, which is used to control a density of a topology view.

[0055] The method for determining a static graph and that for determining a dynamic graph that are applicable to this application are not limited to the examples described above.

[0056] After obtaining the static graphs and the dynamic graphs, the apparatus for performing the method 300 may perform the following steps.

[0057] S320: Generate spatial feature data of the plurality of nodes based on the static graphs and the dynamic graphs.

[0058] For example, the static graphs and the dynamic graphs may be processed by using a graph neural network (graph neural networks, GCN), and a convolution operation may be performed on the static graph and the dynamic graphs separately by using a convolution kernel of the GCN to obtain the spatial feature data.

[0059] S330: Obtain time feature data of the plurality of nodes.

[0060] An execution sequence of S320 and S330 is not limited. Observation data of the plurality of nodes in a historical time range may be processed by using a gated recurrent unit (gate recurrent unit, GRU) in a recursive neural network (recursive neural network, RNN) to obtain the

time feature data.

[0061] The foregoing observation data includes at least one of the following data: meteorological data, network topology data, traffic data, voice data, signaling data, point of interest (point of interest, POI) data, major-event data, and holiday data.

5 **[0062]** Generally, a data amount of the observation data is relatively large, and a time required for performing prediction by directly using the observation data is relatively long. Therefore, before processing the observation data, the observation data may be preprocessed to reduce the time required for prediction.

[0063] For example, the observation data may be sliced (that is, sampled) by using
10 characteristics of the observation data such as temporal proximity and periodicity to determine time slice data from the observation data. A data amount of the time slice data is less than the data amount of the observation data. Then, the time slice data is processed by using the GRU to obtain the time feature data.

[0064] The following describes a method for performing time slicing on traffic data according
15 to this application.

[0065] FIG. 4 shows traffic data within a historical time range (20 days) of a target base station. It can be learned that, the traffic data is apparently periodic. The traffic data may be sliced based on a location of the target time range (that is, a time range in which prediction is to be performed) in one period, and traffic data in a time range that corresponds to the target time range and that is
20 in another period is determined as the time slice data. Because data in corresponding time ranges in different periods is similar, performing prediction based on time slice data obtained based on this feature can reduce the time required for prediction without affecting a prediction effect.

[0066] FIG. 5 shows a method for obtaining time slice data.

[0067] If the target time range is 8:15 to 9:15 ante meridiem (a.m.) on September 1 (Saturday
25 (Sat.)), 2018, a period may be a day (day) or a week (week), and traffic data in a time range that corresponds to the target time range and that is in an adjacent period may be determined as the time slice data.

[0068] For example, if a period (September 1, 2018) in which the target time range (8:15 to 9:15 a.m.) is located is a second period, August 30, 2018 may be regarded as the first period, and
30 8:15 to 9:15 a.m. on August 30, 2018 is a time range that corresponds to the target time range and that is in the first period, and a time range including this corresponding time range may be referred

to as a specific time range. In the example in FIG. 5, the specific time range on August 30, 2018 is 6:30 to 9:30 a.m. Similarly, it may be further determined that 6:30 to 9:30 a.m. on August 30, 2018 is the specific time range. In this example, each specific time range may be referred to as T_d .

[0069] For another example, if a period (August 27, 2018 to September 2, 2018) in which the target time range (8:15 to 9:15 a.m.) is located is a second period, a period from August 20, 2018 to August 26, 2018 may be regarded as the first period, and 8:15 to 9:15 a.m. on August 25, 2018 (Saturday) is a time range that corresponds to the target time range and that is in the first period. A time range including this corresponding time range may be referred to as a specific time range. In the example in FIG. 5, a specific time range on August 20, 2018 to August 26, 2018 is 6:30 to 9:30 a.m. on August 25, 2018 (Saturday). Similarly, it may be further determined that 6:30 to 9:30 a.m. on August 11, 2018 (Saturday) is a specific time range. In this example, each specific time range may be referred to as T_w .

[0070] After the specific time range is determined, observation data in the specific time range is the time slice data.

[0071] Similarly, the time slice data may be further determined from traffic data in a historical time range by using a month (month) or a quarter (quarter) as a period.

[0072] In the example shown in FIG. 5, the time slice data further includes traffic data in a time range adjacent to the target time range, and the adjacent time range is, for example, 5:00 to 8:00 a.m. on September 1, 2018 and 8:00 to 11:00 post meridiem (p.m.) on August 30, 2018. In this example, each adjacent time range may be referred to as T_r .

[0073] It should be noted that, the foregoing several examples about the time slice data may be used independently or in combination.

[0074] If traffic data in a historical time range is data obtained at an interval of 15 minutes, each time slice (T_d , T_w , or T_r) in FIG. 5 includes 12 pieces of data. Three pieces of the data in T_d , three pieces of the data in T_w , and four pieces of the data in T_r may be selected as the time slice data. In this case, there are totally 120 pieces of data (equivalent to data in eight hours) in the time slice data. Traffic data in 20 days includes about 1920 pieces of data in total. Compared with the time slice data, performing prediction by using the time slice data can reduce the time required for prediction without affecting the prediction effect.

[0075] After the time feature data and the spatial feature data are obtained, the apparatus for performing the method 300 may perform the following step.

[0076] S340: Obtain a predicted state of a target node in a target time range based on the spatial feature data and the time feature data, where the target node is any node in the plurality of nodes.

[0077] FIG. 6 shows a prediction method according to this application.

[0078] In FIG. 6, X^t denotes data obtained after time slicing processing is performed on base station observation data existing before a moment t . X^t is input into a temporal model (temporal modeling), and time feature data Y^t is output after processing by the temporal model. For example, the temporal model is a GRU.

[0079] Static graphs and dynamic graphs are input into a spatial model (spatial modeling), and spatial feature data is output after processing by the spatial model. For example, the spatial model

is a GCN, and the time feature data includes $\Theta_{*G_i^t}$. $\Theta_{*G_i^t}$ denotes a result of a convolution operation performed on a graph G_i^t by using a convolution kernel Θ of the GCN. G_i^t is one of the dynamic graphs and the static graphs existing before the moment t .

[0080] Then, Y^t is input into the spatial model. Based on Y^t and $\Theta_{*G_i^t}$, a result Z_i^t combining a temporal feature and a spatial feature may be obtained, as shown in formula (5).

$$Z_i^t = ReLU\left(\Theta_{*G_i^t}(Y^t)\right) \quad (5)$$

[0081] $ReLU$ denotes a rectified linear unit (rectified linear unit).

[0082] A final output result Z^t of the spatial model may be obtained by adding up a plurality of Z_i^t , as shown in formula (6).

$$Z^t = \sum_{i=1}^m Z_i^t \quad (6)$$

[0083] Then, Z^t may be processed by using a fully connected (fully connected, FC) layer to obtain $\hat{x}_{t+k}$, as shown in formula (7).

$$\hat{x}_{t+k} = Z^t W_f + b_f \quad (7)$$

[0084] W_f is a weight parameter of the FC layer, b_f is an offset parameter of the FC layer, and $\hat{x}_{t+k}$ denotes a predicted state of the target base station at a moment $t+k$.

[0085] It can be learned from the above that, in a process of node state prediction, a dynamic

spatial feature obtained from observation data is used in addition to a time feature and a static spatial feature that are obtained from the observation data. This increases accuracy of node state prediction in a network with a dynamic topological relationship.

[0086] In a training process, co-training may be performed on the temporal model, the spatial model, and the FC layer by minimizing a loss function. An optional loss function is shown in formula (8).

$$L(\hat{x}_{t+k}, x_{t+k}) = \sum_{t \in T_{train}} \frac{1}{N} \sum_{i=1}^N \left((\hat{x}_{t+k}(i) - x_{t+k}(i))^2 + \alpha \frac{1}{N} |\hat{x}_{t+k}(i) - x_{t+k}(i)| \right) \quad (8)$$

[0087] $L(\hat{x}_{t+k}, x_{t+k})$ is the loss function, x_{t+k} is a true state of the target base station at the moment $t+k$, $(\hat{x}_{t+k}(i) - x_{t+k}(i))^2$ is a mean square error (mean square error, MSE), $|\hat{x}_{t+k}(i) - x_{t+k}(i)|$ is a mean absolute error (mean absolute error, MAE), α is used to balance weights of the MSE and the MAD, and N denotes a quantity of nodes.

[0088] The following compares the prediction effect of the method 300 with prediction effects of other methods to describe beneficial effects of this application.

[0089] The methods participating in the comparison include historical average (historical average, HA), Prophet (Prophet), Spatio-Temporal Graph Convolutional Network (spatio-temporal graph convolutional network, STGCN), Attention Spatio -Temporal Graph Convolutional Network (attention spatio-temporal graph convolutional network, ASTGCN) and Diffusion Convolutional Recurrent Neural Network (diffusion convolutional recurrent neural network, DCRNN), and evaluation indicators are a root mean square error (root mean square error, RMSE) and an MAE. Smaller values of the RMSE and the MAE indicate a better predictive capability of a method. Comparison results are shown in Table 1.

Table 1

	15 minutes		30 minutes		45 minutes		1 hour	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
HA	108.9	44.2	108.9	44.2	108.9	44.2	108.9	44.2
Prophet	70.3	42.3	72.2	43.0	73.8	43.7	75.2	44.2
STGCN	52.7±1.0	24.7±0.2	66.8±1.6	30.8±0.4	75.3±2.0	34.7±0.6	82.4±2.5	37.9±0.8
ASTGCN	72.4±2.7	28.5±1.2	74.5±2.7	29.9±1.2	78.1±2.6	32.6±1.2	82.1±2.7	35.2±1.3

	15 minutes		30 minutes		45 minutes		1 hour	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
DCRNN	52.6±0.3	24.4±0.1	64.1±0.7	29.8±0.2	71.8±1.0	33.4±0.3	78.5±1.2	36.5±0.5
Method 300	45.6±0.6	21.4±0.2	51.9±0.8	24.9±0.3	56.4±1.3	27.3±0.4	62.3±1.8	30.6±0.5

[0090] The STGCN, ASTGCN, DCRNN, and method 300 are deep learning-based methods. Some differences exist in a result of each operation. Table 1 shows results of 10 predictions separately performed by these methods, and also shows mean values and variances. It can be learned from Table 1 that, the method 300 achieves an optimal prediction effect among the several methods.

[0091] The foregoing describes in detail an example of a method for predicting a node state according to this application. It can be understood that, to implement the foregoing functions, a corresponding apparatus includes a corresponding hardware structure and/or software module for executing each function. A person skilled in the art should easily be aware that, in combination with units and algorithm steps of the examples described in the embodiments disclosed in this specification, this application may be implemented by hardware or a combination of hardware and computer software. Whether a function is performed by hardware or hardware driven by computer software depends on particular applications and design constraints 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 should not be considered that the implementation goes beyond the scope of this application.

[0092] In this application, functional unit division of the apparatus may be performed based on the foregoing method examples. For example, functions may be divided into functional units, or two or more functions may be integrated into one unit. The integrated unit may be implemented in a form of hardware, or may be implemented in a form of a software functional unit. It should be noted that, unit division in this application is an example, and is merely logical function division. During actual implementation, another division manner may be used.

[0093] FIG. 7 is a schematic structural diagram of an apparatus for predicting a node state according to this application. An apparatus 700 includes an input unit 710 and a processing unit 720.

[0094] The processing unit 720 is configured to obtain static graphs and dynamic graphs of a

plurality of nodes in a target network, where the static graphs and the dynamic graphs are all topology views; generate spatial feature data of the plurality of nodes based on the static graphs and the dynamic graphs; obtain time feature data of the plurality of nodes; and obtain a predicted state of a target node in a target time range based on the spatial feature data and the time feature data, where the target node is any node in the plurality of nodes.

[0095] Optionally, the input unit 710 is configured to obtain observation data of the plurality of nodes in a first historical time range, where the first historical time range is a time range before the target time range; and the processing unit 720 is specifically configured to slice the observation data in a time dimension to obtain time slice data, where a data amount of the time slice data is less than a data amount of the observation data; and obtain the time feature data based on the time slice data.

[0096] Optionally, the time slice data includes: observation data in a time range that is in the first historical time range and that is adjacent to the target time range.

[0097] Optionally, the time slice data includes: observation data in a specific time range in the first historical time range, where the specific time range is located in a first period, the target time range is located in a second period, and the specific time range includes a time range that corresponds to the target time range and that is in the first period.

[0098] Optionally, the input unit 710 is configured to obtain observation data of the plurality of nodes in a second historical time range, where the second historical time range is a time range before the target time range; and the processing unit 720 is specifically configured to construct the static graphs and the dynamic graphs based on the observation data.

[0099] Optionally, the observation data includes at least one of the following data: meteorological data, network topology data, traffic data, voice data, signaling data, POI data, major-event data, and holiday data.

[00100] Optionally, the processing unit 720 is further configured to obtain a true state of the target node in the target time range; and train a temporal model and a spatial model based on the true state and the predicted state, where the temporal model is used to output the time feature data based on input observation data, and the spatial model is used to output the spatial feature data based on the static graphs and the dynamic graphs that are input.

[00101] Optionally, the static graphs belong to first-type topology views, the dynamic graphs belong to second-type topology views, and a topological relationship change rate of the first-type

topology views is less than a topological relationship change rate of the second-type topology views.

[00102] Optionally, the target network is a communications network, the static graphs are topology views representing a traffic-based topological relationship of the plurality of nodes, and
5 the dynamic graphs are topology views representing a physical-line-based topological relationship of the plurality of nodes.

[00103] For a specific manner in which the apparatus 700 performs the method for predicting a node state and a beneficial effect incurred, refer to related descriptions in the method embodiments.

[00104] FIG. 8 is a schematic structural diagram of a device for predicting a node state
10 according to this application. A dashed line in FIG. 8 indicates that the unit or the module is optional. A device 800 may be configured to implement the method described in the foregoing method embodiments. The device 800 may be a terminal device, a server, or a chip.

[00105] The device 800 includes one or more processors 801. The one or more processors 801 can support the device 800 in implementing the method in the method embodiments. The processor
15 801 may be a general-purpose processor or a dedicated processor. For example, the processor 801 may be a central processing unit (central processing unit, CPU). The CPU may be configured to control the device 800, execute a software program, and process data of the software program. The device 800 may further include a communications unit 805, configured to implement input (receiving) and/or output (sending) of signals (such as observation data and prediction results).

[00106] For example, the device 800 may be a chip, and the communications unit 805 may be an input circuit and/or an output circuit of the chip, or the communications unit 805 may be a communications interface of the chip. The chip may be used as a component of a terminal device, a network device, or another electronic device.

[00107] For another example, the device 800 may be a terminal device or a server, and the
25 communications unit 805 may be a transceiver of the terminal device or the server, or the communications unit 805 may be a transceiver circuit of the terminal device or the server.

[00108] The device 800 may include one or more memories 802, which store a program 804. The program 804 may be run by the processor 801 to generate an instruction 803, so that the processor 801 performs the method described in the foregoing method embodiments based on the
30 instruction 803. Optionally, the memory 802 may further store data (for example, observation data). Optionally, the processor 801 may further read the data stored in the memory 802, where the data

may be stored at a storage address the same as that of the program 804, or the data may be stored at a storage address different from that of the program 804.

[00109] The processor 801 and the memory 802 may be separately disposed, or may be integrated together, for example, integrated on a system on chip (system on chip, SOC) of a terminal device.

[00110] For a specific manner in which the processor 801 executes the method embodiments, refer to related descriptions in the method embodiments.

[00111] It should be understood that the steps in the foregoing method embodiments may be completed by using a logic circuit in a form of hardware or an instruction in a form of software in the processor 801. The processor 801 may be a CPU, a digital signal processor (digital signal processor, DSP), an application specific integrated circuit (application specific integrated circuit, ASIC), a field programmable gate array (field programmable gate array, FPGA), or another programmable logic device, for example, discrete gates, transistor logic, or discrete hardware components.

[00112] This application further provides a computer program product. When the computer program product is executed by the processor 801, the method according to any method embodiment of this application is implemented.

[00113] The computer program product may be stored in the memory 802, for example, may be the program 804. After processing processes such as preprocessing, compilation, assembly, and linking, the program 804 is finally converted into an executable target file that can be executed by the processor 801.

[00114] This application further provides a computer readable storage medium, which stores a computer program. When the computer program is executed by a computer, the method according to any method embodiment of this application is implemented. The computer program may be a high-level language program or an executable target program.

[00115] For example, the computer readable storage medium is the memory 802. The memory 802 may be a volatile memory or a non-volatile memory, or the memory 802 may include both a volatile memory and a non-volatile memory. The nonvolatile memory may be a read-only memory (read-only memory, ROM), a programmable read-only memory (programmable ROM, PROM), an erasable programmable read-only memory (erasable PROM, EPROM), an electrically erasable programmable read-only memory (electrically EPROM, EEPROM), or a flash memory. The

volatile memory may be a random access memory (random access memory, RAM), used as an external cache. Through example but not limitative description, many forms of RAMs may be used, for example, a static random access memory (static RAM, SRAM), a dynamic random access memory (dynamic RAM, DRAM), a synchronous dynamic random access memory (synchronous
5 DRAM, SDRAM), a double data rate synchronous dynamic random access memory (double data rate SDRAM, DDR SDRAM), an enhanced synchronous dynamic random access memory (enhanced SDRAM, ESDRAM), a synchronous link dynamic random access memory (synchlink DRAM, SLDRAM), and a direct rambus dynamic random access memory (direct rambus RAM, DR RAM).

10 **[00116]** It may be clearly understood by a person skilled in the art that, for ease and brevity of description, for a specific working process and a generated technical effect of the foregoing apparatus and device, refer to a corresponding process and technical effect in the foregoing method embodiments, and details are not described herein again.

[00117] In the several embodiments provided in this application, the disclosed system,
15 apparatus and method may be implemented in other manners. For example, some features of the method embodiments described above may be ignored or not performed. The described apparatus embodiments are merely examples. The unit division is merely logical function division and may be other division in actual implementation. A plurality of units or components may be combined or integrated into another system. In addition, coupling between the units or coupling between the
20 components may be direct coupling or indirect coupling, and the coupling may include an electrical connection, a mechanical connection, or another form of connection.

[00118] Sequence numbers of the foregoing processes do not mean execution sequences in various embodiments of this application. The execution sequences of the processes should be determined based on functions and internal logic of the processes, and should not be construed as
25 any limitation on the implementation processes of the embodiments of this application.

[00119] In addition, the term "and/or" in this specification describes only an association relationship for associated objects and represents that three relationships may exist. For example, A and/or B may represent the following three cases: Only A exists, both A and B exist, and only B exists. In addition, the character "/" in this specification generally indicates an "or" relationship
30 between the associated objects.

[00120] In summary, what is described above is merely example embodiments of the technical

solutions of this application, but is not intended to limit the protection scope of this application. Any modification, equivalent replacement, or improvement made without departing from the principle of this application shall fall within the protection scope of this application.

CLAIMS

What is claimed is:

1. A method for predicting a node state, comprising:

obtaining static graphs and dynamic graphs of a plurality of nodes in a target network, wherein

5 the static graphs and the dynamic graphs are all topology views;

generating spatial feature data of the plurality of nodes based on the static graphs and the dynamic graphs;

obtaining time feature data of the plurality of nodes; and

10 obtaining a predicted state of a target node in a target time range based on the spatial feature data and the time feature data, wherein the target node is any node in the plurality of nodes.

2. The method according to claim 1, wherein the obtaining time feature data of the plurality of nodes comprises:

obtaining observation data of the plurality of nodes in a first historical time range, wherein the first historical time range is a time range before the target time range;

15 slicing the observation data in a time dimension to obtain time slice data, wherein a data amount of the time slice data is less than a data amount of the observation data; and

obtaining the time feature data based on the time slice data.

3. The method according to claim 2, wherein the time slice data comprises:

20 observation data in a time range that is in the first historical time range and that is adjacent to the target time range.

4. The method according to claim 2 or 3, wherein the time slice data comprises:

25 observation data in a specific time range in the first historical time range, wherein the specific time range is located in a first period, the target time range is located in a second period, and the specific time range comprises a time range that corresponds to the target time range and that is in the first period.

5. The method according to any one of claims 1 to 4, wherein the obtaining static graphs and dynamic graphs of a plurality of nodes comprises:

obtaining observation data of the plurality of nodes in a second historical time range, wherein the second historical time range is a time range before the target time range; and

30 constructing the static graphs and the dynamic graphs based on the observation data.

6. The method according to claim 5, wherein the observation data comprises at least one of the following data:

meteorological data, network topology data, traffic data, voice data, signaling data, point of interest POI data, major-event data, and holiday data.

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

obtaining a true state of the target node in the target time range; and

training a temporal model and a spatial model based on the true state and the predicted state, wherein the temporal model is used to output the time feature data based on input observation data, and the spatial model is used to output the spatial feature data based on the static graphs and the
10 dynamic graphs that are input.

8. The method according to any one of claims 1 to 7, wherein the static graphs belong to first-type topology views, the dynamic graphs belong to second-type topology views, and a topological relationship change rate of the first-type topology views is less than a topological relationship change rate of the second-type topology views.

15 9. The method according to any one of claims 1 to 8, wherein the target network is a communications network, the static graphs are topology views representing a traffic-based topological relationship of the plurality of nodes, and the dynamic graphs are topology views representing a physical-line-based topological relationship of the plurality of nodes.

10. An apparatus for predicting a node state, comprising a processing unit configured to:

20 obtain static graphs and dynamic graphs of a plurality of nodes in a target network, wherein the static graphs and the dynamic graphs are all topology views;

generate spatial feature data of the plurality of nodes based on the static graphs and the dynamic graphs;

obtain time feature data of the plurality of nodes; and

25 obtain a predicted state of a target node in a target time range based on the spatial feature data and the time feature data, wherein the target node is any node in the plurality of nodes.

11. The apparatus according to claim 10, wherein the apparatus further comprises an input unit, wherein

the input unit is configured to obtain observation data of the plurality of nodes in a first
30 historical time range, wherein the first historical time range is a time range before the target time range;

the processing unit is specifically configured to slice the observation data in a time dimension to obtain time slice data, wherein a data amount of the time slice data is less than a data amount of the observation data; and obtain the time feature data based on the time slice data.

12. The apparatus according to claim 11, wherein the time slice data comprises:

5 observation data in a time range that is in the first historical time range and that is adjacent to the target time range.

13. The apparatus according to claim 11 or 12, wherein the time slice data comprises:

observation data in a specific time range in the first historical time range, wherein the specific time range is located in a first period, the target time range is located in a second period, and the
10 specific time range comprises a time range that corresponds to the target time range and that is in the first period.

14. The apparatus according to any one of claims 10 to 13, wherein the apparatus further comprises an input unit, wherein

the input unit is configured to obtain observation data of the plurality of nodes in a second
15 historical time range, wherein the second historical time range is a time range before the target time range; and

the processing unit is specifically configured to construct the static graphs and the dynamic graphs based on the observation data.

15. The apparatus according to claim 14, wherein the observation data comprises at least one
20 of the following data:

meteorological data, network topology data, traffic data, voice data, signaling data, point of interest POI data, major-event data, and holiday data.

16. The apparatus according to any one of claims 10 to 15, wherein the processing unit is further configured to:

25 obtain a true state of the target node in the target time range; and

train a temporal model and a spatial model based on the true state and the predicted state, wherein the temporal model is used to output the time feature data based on input observation data, and the spatial model is used to output the spatial feature data based on the static graphs and the dynamic graphs that are input.

30 17. The apparatus according to any one of claims 10 to 16, wherein the static graphs belong to first-type topology views, the dynamic graphs belong to second-type topology views, and a

topological relationship change rate of the first-type topology views is less than a topological relationship change rate of the second-type topology views.

18. The apparatus according to any one of claims 10 to 17, wherein the target network is a communications network, the static graphs are topology views representing a traffic-based
5 topological relationship of the plurality of nodes, and the dynamic graphs are topology views representing a physical-line-based topological relationship of the plurality of nodes.

19. A device for predicting a node state, wherein the device comprises a processor and a memory, the memory is configured to store a computer program, and the processor is configured to invoke and run the computer program from the memory, so as to enable the device to perform
10 the method according to any one of claims 1 to 9.

20. A computer readable storage medium, wherein the computer readable storage medium stores a computer program, and when the computer program is executed by a processor, the processor is enabled to perform the method according to any one of claims 1 to 9.