

WHAT IS CLAIMED IS:

1. A neural network system for processing execution data~~a first calculation processing graph~~, the neural network system comprising:

a first compiler to generate first execution data by compiling a first subgraph, the first subgraph being included in a first calculation processing graph;

a first resource to receive the first execution data, and to generate a first calculation result obtained by performing at least one calculation operation on the first execution data~~corresponding to the first calculation processing graph~~; and

a task manager to calculate a predicted first polling time taken for the first resource to perform the at least one calculation operation and to poll the first calculation result from the first resource after ~~based on the predicted~~ calculated first polling time,

wherein the first resource is to output the first calculation result in response to the polling by the task manager, and

wherein the task manager is to assign information about the first subgraph into a predetermined equation to calculate the predicted first polling time.

2. The neural network system as claimed in claim 1, wherein

~~the first calculation processing graph includes a first subgraph to be processed by the first resource,~~

~~the neural network system further includes a first compiler to generate first execution data by compiling the first subgraph to be driven by the first resource, and~~

the task manager is to output the first execution data to the first resource to perform a first calculation operation corresponding to the first subgraph and to poll the first calculation result from the first resource after the predicted first polling time elapses from a time when the first execution data is output.

3. The neural network system as claimed in claim 2, wherein the first calculation processing graph further includes a second subgraph, the neural network system further includes:

a second resource to generate a second calculation result obtained by performing a second calculation operation corresponding to the second subgraph based on the first calculation result and second execution data; and

a second compiler to generate the second execution data by compiling the first subgraph to be driven by the second resource, and

the task manager further to calculate a predicted second polling time for which the second resource performs the second calculation operation and to poll the second calculation result from the second resource after the predicted first polling time and the predicted second polling time elapse from a time when the first execution data is output to the first resource.

4. The neural network system as claimed in claim 3, further comprising a model analyzer to receive a first input model, analyze the first input model to generate the first calculation processing graph, and divide the first calculation processing graph into a plurality of subgraphs comprising the first subgraph and the second subgraph.

5. The neural network system as claimed in claim 4, wherein the model analyzer comprises:

a model parser to receive the first input model and extract model information comprising size information about the input model from the first input model;

a model builder to analyze the first input model based on the model information to generate the first calculation processing graph; and

a model optimizer to perform an optimization operation on the first calculation processing graph.

6. The neural network system as claimed in claim 4 or claim 5, wherein

the model analyzer is to receive a second input model and analyze the second input model to generate a second calculation processing graph comprising a third subgraph,

the first resource is configured to further generate a third calculation result obtained by performing a third calculation operation corresponding to the third subgraph, and

the task manager is configured to calculate a predicted third polling time for which the first resource performs the third calculation operation and to poll the third calculation result from the first resource based on the predicted~~calculated~~ third polling time.

7. The neural network system as claimed in claim 6, wherein the task manager is to poll the third calculation result from the first resource after the predicted third polling time elapses from a time when the second resource completes the second calculation operation.

8. The neural network system as claimed in claim 6, wherein the task manager is to poll the third calculation result from the first resource after the predicted third polling time elapses from a time when the first resource completes the first calculation operation.

~~9. The neural network system as claimed in any one of claims 2 to 8, wherein the task manager is to assign information about the first subgraph into a predetermined equation to calculate the first polling time.~~

~~109.~~ The neural network system as claimed in any one of claims 1 to 98, wherein the information about the first subgraph includes at least one of size information about an input model corresponding to the first calculation processing graph, size information about the first subgraph, size information about the first execution data, and a compiling time for which the first compiler compiles the first subgraph.

~~11~~10. The neural network system as claimed in any one of claims ~~2-1~~ to ~~10~~9, wherein the task manager calculates the predicted first polling time based on a calculation processing ability of the first resource.

~~12~~11. The neural network system as claimed in any one of claims 2 to ~~11~~10, wherein

the first calculation operation includes n (where n is an integer equal to or greater than zero) number of first sub-calculations and m (where m is an integer equal to or greater than zero) number of second sub-calculations,

a first time is taken to perform a first sub-calculation by the first resource, and a second time is taken to perform the second sub-calculation by the first resource, and

the task manager is to calculate the predicted first polling time as a sum of n times the first time and m times the second time.

~~13~~12. The neural network system as claimed in ~~claim 2 or claim 3~~any one of claims 1 to 3, wherein

the first calculation processing graph includes a plurality of subgraphs, and

the neural network system further includes a model keeper to store a polling timetable including information about the plurality of subgraphs and information about a predicted polling time corresponding to each of the plurality of subgraphs.

~~14~~13. The neural network system as claimed in claim ~~13~~12, wherein

the task manager is to determine whether the polling timetable stores the predicted first polling time corresponding to the first subgraph before calculating a predicted polling time corresponding to the first subgraph, and

when the predicted first polling time is stored in the polling timetable, the task manager is to perform polling based on the predicted first polling time stored in the polling timetable.

1514. The neural network system as claimed in claim 1312 or claim 1413, wherein, when the predicted first polling time is not stored, the task manager updates the polling timetable with the predicted~~calculated~~ first polling time.

1615. The neural network system as claimed in any one of claims 1312 to 1514, wherein the polling timetable further includes information about a resource characteristic where calculation corresponding to each of the plurality of subgraphs is performed.

1716. A method of processing a ~~neural network model using a calculation~~ processing graph including a first subgraph~~execution data~~, the method comprising:

calculating, by a task manager, a predicted first polling time taken to perform a first calculation operation corresponding to ~~the a first subgraph, based on information about the first subgraph~~ the first subgraph being included in a first calculation processing graph;

generating, by a first compiler, first execution data by compiling the first subgraph ~~to be driven by a first resource~~;

outputting, by the task manager, the first execution data to ~~the a first resource; and~~ generating, by the first resource, a first calculation result obtained by performing the first calculation operation on the first execution data;

polling, by the task manager, ~~a the~~ first calculation result corresponding to the first execution data from the first resource after the predicted first polling time ~~elapses from a time when the first execution data is output to the first resource; and~~

outputting, by the first resource, the first calculation result in response to the polling by the task manager.

wherein the calculating the predicted first polling time comprises: assigning, by the task manager, information about the first subgraph into a predetermined equation to calculate the predicted first polling time.

1817. A computer-based neural network system for processing execution data ~~finding a calculation processing graph corresponding to a structure of data to perform at least one calculation operation corresponding to the calculation processing graph,~~ the computer-based neural network system comprising:

an input device to receive input data;

a memory to program information comprising information about a model analyzer to find a calculation processing graph from the input data and information about a task manager to generate a calculation result corresponding to the calculation processing graph based on the calculation processing graph;

~~— a processor to process the input data based on the program information~~

a compiler to generate the execution data by compiling a subgraph, the subgraph being included in the calculation processing graph; and

a resource to receive the execution data from the task manager, and generate the calculation result obtained by performing the at least one calculation operation on the execution data~~corresponding to the calculation processing graph,~~

wherein the ~~task manager~~processor is to calculate a predicted polling time taken to perform the at least one calculation operation ~~based on program information about the task manager~~ and to poll the calculation result from the resource ~~based on~~after the predicted polling time,

wherein the resource is to output the calculation result in response to the polling by the task manager, and

wherein the task manager is to assign information about the subgraph into a predetermined equation to calculate the predicted polling time.

1918. The computer-based neural network system as claimed in claim 1817, wherein ~~the calculation processing graph includes a subgraph processed by the resource,~~ the computer-based neural network system further comprises:

~~— a compiler to generate execution data by compiling the subgraph to be driven by the resource, and~~

——the ~~task manager~~processor is to output the execution data to the resource, for performing a calculation operation corresponding to the subgraph based on the program information about the task manager and to poll the calculation result from the resource after the predicted polling time elapses from a time when the execution data is output.

——20.——The computer-based neural network system as claimed in claim 19, wherein the processor is to assign information about the subgraph into a predetermined equation to calculate the polling time.