

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7680552号
(P7680552)

(45)発行日 令和7年5月20日(2025.5.20)

(24)登録日 令和7年5月12日(2025.5.12)

(51)国際特許分類 F I
G 0 6 F 9/50 (2006.01) G 0 6 F 9/50 1 5 0 Z

請求項の数 12 (全50頁)

(21)出願番号	特願2023-548262(P2023-548262)	(73)特許権者	518389015 中国移动通信有限公司研究院 China Mobile Communication Co., Ltd Research Institute
(86)(22)出願日	令和4年2月8日(2022.2.8)		中華人民共和国100053北京市西城区宣武門西大街32号
(65)公表番号	特表2024-506073(P2024-506073 A)		32 Xuanwumen West Street, Xicheng District, Beijing 100053, China
(43)公表日	令和6年2月8日(2024.2.8)		
(86)国際出願番号	PCT/CN2022/075450	(73)特許権者	507142144 中国移动通信集团有限公司 CHINA MOBILE COMMUNICATIONS GROUP CO.,
(87)国際公開番号	WO2022/171066		最終頁に続く
(87)国際公開日	令和4年8月18日(2022.8.18)		
審査請求日	令和5年9月29日(2023.9.29)		
(31)優先権主張番号	202110184998.5		
(32)優先日	令和3年2月10日(2021.2.10)		
(33)優先権主張国・地域又は機関	中国(CN)		

(54)【発明の名称】 モノのインターネット機器ベースのタスク割り当て方法、ネットワーク訓練方法及びその装置

(57)【特許請求の範囲】

【請求項1】

電子機器によって実行される、モノのインターネット機器ベースのネットワーク訓練方法であって、

前記ネットワーク訓練方法は、訓練データ集合を決定し、前記訓練データ集合に基づいて第1ネットワークを訓練することを含み、前記訓練データ集合は、少なくとも1つのタスク割り当てポリシー及び対応する実際のシステム性能を含み、1つの実際のシステム性能は、対応するタスク割り当てポリシーに基づいて実際に実行することによって得られ、前記第1ネットワークは、タスク割り当てポリシーに対応するシステム性能を予測するために使用され、

前記ネットワーク訓練方法は、

処理対象タスクに対応する計算グラフ、及び1つ又は複数のモノのインターネット機器に対応するリソースグラフを決定し、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成することを更に含み、

前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成することは、

前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成することを含み、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表し、前記タスク割り当てポリシーは、対応するリソースグラフの少なくとも1つのノードを前記計算グラフの各ノードに割り

当てるために使用され、リソースサブグラフ内の1つのノードは、1つ又は複数のモノのインターネット機器の能力及びノ又はリソースの少なくとも一部を表し、リソースサブグラフ内の隣接する2つのノードのエッジは、1つ又は複数のモノのインターネット機器の能力及びノ又はリソースの少なくとも一部の間の関係を表す、ネットワーク訓練方法。

【請求項2】

前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成することは、

前記計算グラフ内の第1ノードを決定することであって、前記第1ノードは、リソース需要が最も大きいノードであり、前記リソース需要が最も大きいノードは、計算リソース、ストレージリソース及び通信リソースのうちの少なくとも1つのリソースの需要が最も大きいノードである、ことと、

10

前記リソースグラフ内の少なくとも1つの第2ノードを決定することであって、前記少なくとも1つの第2ノードは、前記第1ノードのリソース需要を満たすノードである、ことと、

各第2ノードに基づいて、1つのリソースサブグラフを決定することとを含む、請求項1に記載のネットワーク訓練方法。

【請求項3】

前記第1ネットワークを訓練することは、

少なくとも1つのタスク割り当てポリシーに対応する予測のシステム性能及び実際のシステム性能に基づいて、前記第1ネットワークを訓練することを含み、

20

計算グラフ及び各リソースサブグラフに基づいて、第1ネットワークを用いて各リソースサブグラフに対応する予測のシステム性能を得る、請求項1に記載のネットワーク訓練方法。

【請求項4】

前記第1ネットワークを用いて各リソースサブグラフに対応する予測のシステム性能を得ることは、

前記第1ネットワークの特徴抽出モジュールを用いて前記計算グラフの特徴を抽出し、第1特徴集合を得ることと、

前記特徴抽出モジュールを用いて、前記少なくとも1つのリソースサブグラフの特徴をそれぞれ抽出し、少なくとも1つの第2特徴集合を得ることと、

30

前記第1特徴集合、各第2特徴集合及び前記第1ネットワークの予測モジュールに基づいて、各リソースサブグラフに対応する予測のシステム性能を得ることとを含む、請求項3に記載のネットワーク訓練方法。

【請求項5】

前記第1特徴集合、各第2特徴集合及び前記第1ネットワークの予測モジュールに基づいて、各リソースサブグラフに対応する予測のシステム性能を得ることは、

前記第1特徴集合及び各第2特徴集合に基づいて、少なくとも1つの第3特徴集合を得ることであって、各第3特徴集合は、前記第1特徴集合及び各第2特徴集合を含む、ことと、

各第3特徴集合及び前記予測モジュールに基づいて、各リソースサブグラフに対応する予測データを得、各リソースサブグラフに対応する予測データに基づいて、各リソースサブグラフに対応する予測のシステム性能を得ることと

40

を含む、請求項4に記載のネットワーク訓練方法。

【請求項6】

前記第1ネットワークを訓練することは、

各タスク割り当てポリシーに対応する予測のシステム性能及び実際のシステム性能に基づいて、前記特徴抽出モジュールと前記予測モジュールを訓練することを含む、請求項5に記載のネットワーク訓練方法。

【請求項7】

前記特徴抽出モジュールと前記予測モジュールを訓練することは、

50

各タスク割り当てポリシーに対応する予測のシステム性能と実際のシステム性能の間の誤差を逆伝播し、予測のシステム性能と実際のシステム性能の間の誤差が所定条件を満たすまで、勾配降下アルゴリズムを利用して第1ネットワークの特徴抽出モジュールと予測モジュールのネットワークパラメータを更新することを含む、請求項6に記載のネットワーク訓練方法。

【請求項8】

前記ネットワーク訓練方法は、前記訓練データ集合を更新することを更に含み、

前記訓練データ集合を更新することは、

計算グラフ及びリソースグラフに基づいて、ヒューリスティック法、グラフ探索法、グラフ最適化方法及びサブグラフマッチング方法のうち少なくとも1つの方法を用いて、少なくとも1つのリソースサブグラフを生成し、各リソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、各リソースサブグラフに対応する実際のシステム性能を得、前記計算グラフ、各リソースサブグラフ及び対応する実際のシステム性能を前記訓練データ集合に追加すること、

10

計算グラフ及びリソースグラフに基づいて、ヒューリスティック法、グラフ探索法、グラフ最適化方法及びサブグラフマッチング方法のうち少なくとも1つの方法を用いて、少なくとも1つのリソースサブグラフを生成し、第1ネットワークを介して各リソースサブグラフに対応する予測のシステム性能を得、前記少なくとも1つのリソースサブグラフから予測のシステム性能が最も良いリソースサブグラフを選択し、前記予測のシステム性能が最も良いリソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、実際のシステム性能を得、前記計算グラフ、前記予測のシステム性能が最も良いリソースサブグラフ及び対応する実際のシステム性能を前記訓練データ集合に追加すること、

20

計算グラフ及びリソースグラフに基づいて、ランダムウォーク方法を用いて、少なくとも1つのリソースサブグラフを生成し、各リソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、実際のシステム性能を得、前記計算グラフ、少なくとも1つのリソースサブグラフ及び実際のシステム性能を前記訓練データ集合に追加すること

のうちの少なくとも1つを含む、請求項1に記載のネットワーク訓練方法。

【請求項9】

電子機器によって実行される、モノのインターネット機器ベースのタスク割り当て方法であって、前記タスク割り当て方法は、

30

処理対象タスクに対応する計算グラフ、及び1つ又は複数のモノのインターネット機器に対応するリソースグラフを決定することと、

前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成することと、

前記少なくとも1つのタスク割り当てポリシーを第1ネットワークに入力し、各タスク割り当てポリシーに対応する予測のシステム性能を取得することと、

予測のシステム性能が最も良いタスク割り当てポリシーを決定し、決定されたタスク割り当てポリシーに基づいてタスク割り当てを行うことと

40

を含み、

前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成することは、

前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成することを含み、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、又は1つのタスク割り当てポリシーを表し、前記タスク割り当てポリシーは、対応するリソースグラフの少なくとも1つのノードを前記計算グラフの各ノードに割り当てるために使用され、リソースサブグラフ内の1つのノードは、1つ又は複数のモノのインターネット機器の能力及び/又はリソースの少なくとも一部を表し、リソースサブグラフ内の隣接する2つのノードのエッジは、1つ又は複数のモノのインターネット機器の能

50

力及びノ又はリソースの少なくとも一部の間の関係を表す、タスク割り当て方法。

【請求項 10】

前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成することは、

前記計算グラフ内の第1ノードを決定することであって、前記第1ノードは、リソース需要が最も大きいノードであり、前記リソース需要が最も大きいノードは、計算リソース、ストレージリソース及び通信リソースのうちの少なくとも1つのリソースの需要が最も大きいノードである、ことと、

前記リソースグラフ内の少なくとも1つの第2ノードを決定することであって、前記少なくとも1つの第2ノードは、前記第1ノードのリソース需要を満たすノードである、ことと、

各第2ノードに基づいて、1つのリソースサブグラフを決定することと

を含む、請求項9に記載のタスク割り当て方法。

【請求項 11】

モノのインターネット機器ベースのネットワーク訓練装置であって、前記ネットワーク訓練装置は、第1決定ユニットと訓練ユニットとを備え、

前記第1決定ユニットは、訓練データ集合を決定するように構成され、前記訓練データ集合は、少なくとも1つのタスク割り当てポリシー及び対応する実際のシステム性能を含み、1つの実際のシステム性能は、対応するタスク割り当てポリシーに基づいて実際に実行することによって得られ、

前記訓練ユニットは、前記訓練データ集合に基づいて第1ネットワークを訓練するように構成され、前記第1ネットワークは、タスク割り当てポリシーの性能を予測するために使用され、

前記第1決定ユニットはさらに、処理対象タスクに対応する計算グラフ、及び1つ又は複数のモノのインターネット機器に対応するリソースグラフを決定し、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成するように構成され、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表し、前記タスク割り当てポリシーは、対応するリソースグラフの少なくとも1つのノードを前記計算グラフの各ノードに割り当てるために使用され、リソースサブグラフ内の1つのノードは、1つ又は複数のモノのインターネット機器の能力及びノ又はリソースの少なくとも一部を表し、リソースサブグラフ内の隣接する2つのノードのエッジは、1つ又は複数のモノのインターネット機器の能力及びノ又はリソースの少なくとも一部の間の関係を表す、ネットワーク訓練装置。

【請求項 12】

モノのインターネット機器ベースのタスク割り当て装置であって、前記タスク割り当て装置は、第2決定ユニットと第2生成ユニットと予測ユニットとタスク割り当てユニットとを備え、

前記第2決定ユニットは、処理対象タスクに対応する計算グラフ、及び1つ又は複数のモノのインターネット機器に対応するリソースグラフを決定するように構成され、

前記第2生成ユニットは、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成するように構成され、

前記予測ユニットは、前記少なくとも1つのタスク割り当てポリシーを第1ネットワークに入力し、各タスク割り当てポリシーに対応する予測のシステム性能を取得するように構成され、

前記タスク割り当てユニットは、予測のシステム性能が最も良いタスク割り当てポリシーを決定し、決定されたタスク割り当てポリシーに基づいてタスク割り当てを行うように構成され、

前記第2生成ユニットはさらに、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成するように構成され、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、又は1つのタスク割り当てポリシーを表

10

20

30

40

50

し、前記タスク割り当てポリシーは、対応するリソースグラフの少なくとも1つのノードを前記計算グラフの各ノードに割り当てるために使用され、リソースサブグラフ内の1つのノードは、1つ又は複数のモノのインターネット機器の能力及び/又はリソースの少なくとも一部を表し、リソースサブグラフ内の隣接する2つのノードのエッジは、1つ又は複数のモノのインターネット機器の能力及び/又はリソースの少なくとも一部の間を表す、タスク割り当て装置。

【発明の詳細な説明】

【技術分野】

【0001】

(関連出願への相互参照)

本願は、2021年02月10日に中国特許局に提出された、出願番号が202110184998.5である中国特許出願に基づいて提出されるものであり、当該中国特許出願の優先権を主張し、当該中国特許出願の全ての内容は、引用によって本願に組み込まれる。

【0002】

本開示は、モノのインターネット(IoT: Internet of Things)分野に関し、具体的には、1つ又は複数のモノのインターネット機器ベースのタスク割り当て方法、ネットワーク訓練方法及びその装置に関する。

【背景技術】

【0003】

5Gモバイルネットワークと深層学習技術の画期的な進展に伴い、人工知能ベースのモノのインターネットアプリケーションとサービスも新たな機会と挑戦を迎えることになり、あらゆる物事のインターネットが関連する主要な技術トレンドであることは間違いない。クラウドコンピューティングは、計算集約的な深層学習タスクの計算能力とストレージリソースの要件を満たすことができるが、自動運転、仮想現実(VR: Virtual Reality)、増強現実(AR: Augmented Reality)などの遅延、信頼性、プライバシーなどに敏感なモノのインターネットの場面に適用できず、単一のモノのインターネット機器上のリソースは非常に限られているため、相互接続された複数の異機種モノのインターネット機器上で機器間の連携を実行できる分散エッジコンピューティングは、効果的な解決策となる可能性があり、異機種機器間でのインテリジェントな計算タスク割り当て方法がその実現の鍵となる。

【0004】

現在、エッジ深層学習システムでは、深層学習モデルの分散訓練と推論は、主に、モデル分割に基づくレイヤースケジューリングアルゴリズムによって実現されており、モデルの特定のレイヤーは、エッジ側に割り当てられ、残りのレイヤーは、クラウドセンターに割り当てられ、エッジサーバは、主に、低レベルのデータを処理するために使用され、クラウドサーバは、主に、高レベルのデータを処理するために使用される。このようなタスク割り当てポリシーは、基礎となる深層学習アルゴリズムの割り当てを伴わないため、タスクのスケジューリングとリソースの最適化の効果が制限される。

【発明の概要】

【課題を解決するための手段】

【0005】

本開示の実施例は、1つ又は複数のモノのインターネット機器ベースのタスク割り当て方法、ネットワーク訓練方法及びその装置を提供する。

【0006】

本発明の実施例の技術的解決策は、以下のように実現される。

【0007】

第1態様において、本開示の実施例は、1つ又は複数のモノのインターネット機器ベースのネットワーク訓練方法を提供し、前記方法は、訓練データ集合を決定し、前記訓練データ集合に基づいて第1ネットワークを訓練することを含み、前記訓練データ集合は、少

10

20

30

40

50

なくとも1つのタスク割り当てポリシー及び対応する実際の性能を含み、1つの実際の性能は、対応するタスク割り当てポリシーに基づいて実際に実行することによって得られ、前記第1ネットワークは、タスク割り当てポリシーの性能を予測するために使用される。

【0008】

本開示のいくつかの任意の実施例において、前記方法は、処理対象タスクに対応する計算グラフ、及び1つ又は複数のモノのインターネット機器に対応するリソースグラフを決定し、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成することを更に含む。

【0009】

本開示のいくつかの任意の実施例において、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成することは、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成することを含み、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表し、前記タスク割り当てポリシーは、対応するリソースグラフの少なくとも1つのノードを前記計算グラフの各ノードに割り当てるために使用され、リソースサブグラフ内の1つのノードは、1つ又は複数のモノのインターネット機器の能力及びノ又はリソースの少なくとも一部を表し、リソースサブグラフ内の隣接する2つのノードのエッジは、1つ又は複数のモノのインターネット機器の能力及びノ又はリソースの少なくとも一部の間 の関係を表す。

【0010】

本開示のいくつかの任意の実施例において、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成することは、

前記計算グラフ内の第1ノードを決定することであって、前記第1ノードは、リソース需要が最も大きいノードである、ことと、

前記リソースグラフ内の少なくとも1つの第2ノードを決定することであって、前記少なくとも1つの第2ノードは、前記第1ノードのリソース需要を満たすノードである、ことと、

各第2ノードに基づいて、1つのリソースサブグラフを決定することであって、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表す、ことと、を含む。

【0011】

本開示のいくつかの任意の実施例において、前記第1ネットワークを訓練することは、少なくとも1つのタスク割り当てポリシーの予測性能及び実際の性能に基づいて、前記第1ネットワークを訓練することを含み。

【0012】

本開示のいくつかの任意の実施例において、少なくとも1つのタスク割り当てポリシーの予測性能を得ることは、

計算グラフ及び各リソースサブグラフに基づいて、第1ネットワークを用いて各リソースサブグラフに対応する予測性能を得ることを含む。

【0013】

本開示のいくつかの任意の実施例において、前記第1ネットワークを用いて各リソースサブグラフに対応する予測性能を得ることは、

前記第1ネットワークの特徴抽出モジュールを用いて前記計算グラフの特徴を抽出し、第1特徴集合を得ることと、

前記特徴抽出モジュールを用いて、前記少なくとも1つのリソースサブグラフの特徴をそれぞれ抽出し、少なくとも1つの第2特徴集合を得ることと、

前記第1特徴集合、各第2特徴集合及び前記第1ネットワークの予測モジュールに基づいて、各リソースサブグラフに対応する予測性能を得ることと、を含む。

【0014】

本開示のいくつかの任意の実施例において、前記第1特徴集合、各第2特徴集合及び前

10

20

30

40

50

記第 1 ネットワークの予測モジュールに基づいて、各リソースサブグラフに対応する予測性能を得ることは、

前記第 1 特徴集合及び各第 2 特徴集合に基づいて、少なくとも 1 つの第 3 特徴集合を得ることであって、各第 3 特徴集合は、前記第 1 特徴集合及び各第 2 特徴集合を含む、ことと、

各第 3 特徴集合及び前記予測モジュールに基づいて、各リソースサブグラフに対応する予測データを得、各リソースサブグラフに対応する予測データに基づいて、各リソースサブグラフに対応する予測性能を得ることと、を含む。

【 0 0 1 5 】

本開示のいくつかの任意の実施例において、前記予測データは、
前記処理対象タスクを実行するための予測実行時間長と、
前記処理対象タスクを実行するための予測エネルギー消費量と、
前記処理対象タスクを実行するための予測信頼性と、のうちの少なくとも 1 つを含む。

10

【 0 0 1 6 】

本開示のいくつかの任意の実施例において、前記各リソースサブグラフに対応する予測データに基づいて、各リソースサブグラフに対応する予測性能を得ることは、所定重みに従って、各リソースサブグラフに対応する予測データに対して重み付け処理を行い、各リソースサブグラフに対応する予測性能を得ることを含む。

【 0 0 1 7 】

本開示のいくつかの任意の実施例において、前記第 1 ネットワークを訓練することは、各タスク割り当てポリシーの予測性能及び実際の性能に基づいて、前記特徴抽出モジュールと前記予測モジュールを訓練することを含む。

20

【 0 0 1 8 】

本開示のいくつかの任意の実施例において、前記特徴抽出モジュールと前記予測モジュールを訓練することは、

各タスク割り当てポリシーの予測性能と実際の性能の間の誤差を逆伝播し、予測性能と実際の性能の間の誤差が所定条件を満たすまで、勾配降下アルゴリズムを利用して第 1 ネットワークの特徴抽出モジュールと予測モジュールのネットワークパラメータを更新することを含む。

【 0 0 1 9 】

本開示のいくつかの任意の実施例において、前記方法は、前記訓練データ集合を更新することを更に含み、更新された前記訓練データ集合は、前記第 1 ネットワークを更新するために使用される。

30

【 0 0 2 0 】

本開示のいくつかの任意の実施例において、前記訓練データ集合を更新することは、
計算グラフ及びリソースグラフに基づいて、ヒューリスティック法、グラフ探索法、グラフ最適化方法及びサブグラフマッチング方法のうちの少なくとも 1 つの方法を用いて、少なくとも 1 つのリソースサブグラフを生成し、各リソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、各リソースサブグラフに対応する実際の性能を得、前記計算グラフ、各リソースサブグラフ及び対応する実際の性能を前記訓練データ集合に追加すること、

40

計算グラフ及びリソースグラフに基づいて、ヒューリスティック法、グラフ探索法、グラフ最適化方法及びサブグラフマッチング方法のうちの少なくとも 1 つの方法を用いて、少なくとも 1 つのリソースサブグラフを生成し、第 1 ネットワークを介して各リソースサブグラフに対応する予測性能を得、前記少なくとも 1 つのリソースサブグラフから予測性能が最も良いリソースサブグラフを選択し、前記予測性能が最も良いリソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、実際の性能を得、前記計算グラフ、前記予測性能が最も良いリソースサブグラフ及び対応する実際の性能を前記訓練データ集合に追加すること、

計算グラフ及びリソースグラフに基づいて、ランダムウォーク方法を用いて、少なくと

50

も1つのリソースサブグラフを生成し、各リソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、実際の性能を得、前記計算グラフ、少なくとも1つのリソースサブグラフ及び実際の性能を前記訓練データ集合に追加すること、のうちの少なくとも1つを含む。

【0021】

第2態様において、本開示の実施例は更に、モノのインターネット機器ベースのタスク割り当て方法を提供し、前記方法は、

処理対象タスクに対応する計算グラフ、及び1つ又は複数のモノのインターネット機器に対応するリソースグラフを決定することと、

前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成することと、

前記少なくとも1つのタスク割り当てポリシーを第1ネットワークに入力し、各タスク割り当てポリシーに対応する予測性能を取得することと、

予測性能が最も良いタスク割り当てポリシーを決定し、決定されたタスク割り当てポリシーに基づいてタスク割り当てを行うことと、を含む。

【0022】

本開示のいくつかの任意の実施例において、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成することは、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成することを含み、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表し、前記タスク割り当てポリシーは、対応するリソースグラフの少なくとも1つのノードを前記計算グラフの各ノードに割り当てるために使用され、リソースサブグラフ内の1つのノードは、1つ又は複数のモノのインターネット機器の能力及び/又はリソースの少なくとも一部を表し、リソースサブグラフ内の隣接する2つのノードのエッジは、1つ又は複数のモノのインターネット機器の能力及び/又はリソースの少なくとも一部の間関係を表す。

【0023】

本開示のいくつかの任意の実施例において、前記前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成することは、前記計算グラフ内の第1ノードを決定することであって、前記第1ノードは、リソース需要が最も大きいノードである、ことと、

前記リソースグラフ内の少なくとも1つの第2ノードを決定することであって、前記少なくとも1つの第2ノードは、前記第1ノードのリソース需要を満たすノードである、ことと、

各第2ノードに基づいて、1つのリソースサブグラフを決定することであって、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表す、ことと、を含む。

【0024】

本開示のいくつかの任意の実施例において、前記第1ネットワークは、本開示の実施例の第1態様に記載の方法によって最適化される。

【0025】

本開示のいくつかの任意の実施例において、前記各タスク割り当てポリシーに対応する予測性能を得ることは、計算グラフ及び各リソースサブグラフに基づいて、第1ネットワークを用いて各リソースサブグラフに対応する予測性能を得ることを含む。

【0026】

本開示のいくつかの任意の実施例において、前記第1ネットワークを用いて各リソースサブグラフに対応する予測性能を得ることは、

前記第1ネットワークの特徴抽出モジュールを用いて前記計算グラフの特徴を抽出し、第1特徴集合を得ることと、

前記特徴抽出モジュールを用いて、前記少なくとも1つのリソースサブグラフの特徴を

10

20

30

40

50

それぞれ抽出し、少なくとも1つの第2特徴集合を得ることと、

前記第1特徴集合、各第2特徴集合及び前記第1ネットワークの予測モジュールに基づいて、各リソースサブグラフに対応する予測性能を得ることと、を含む。

【0027】

本開示のいくつかの任意の実施例において、前記第1特徴集合、各第2特徴集合及び前記第1ネットワークの予測モジュールに基づいて、各リソースサブグラフに対応する予測性能を得ることは、

前記第1特徴集合及び各第2特徴集合に基づいて、少なくとも1つの第3特徴集合を得ることであって、各第3特徴集合は、前記第1特徴集合及び各第2特徴集合を含む、ことと、

各第3特徴集合及び前記予測モジュールに基づいて、各リソースサブグラフに対応する予測データを得、各リソースサブグラフに対応する予測データに基づいて、各リソースサブグラフに対応する予測性能を得ることと、を含む。

【0028】

本開示のいくつかの任意の実施例において、前記予測データは、

前記処理対象タスクを実行するための予測実行時間長と、

前記処理対象タスクを実行するための予測エネルギー消費量と、

前記処理対象タスクを実行するための予測信頼性と、のうちの少なくとも1つを含む。

【0029】

本開示のいくつかの任意の実施例において、前記各リソースサブグラフに対応する予測データに基づいて、各リソースサブグラフに対応する予測性能を得ることは、所定重みに従って、各リソースサブグラフに対応する予測データに対して重み付け処理を行い、各リソースサブグラフに対応する予測性能を得ることを含む。

【0030】

本開示のいくつかの任意の実施例において、前記方法は、タスク割り当てを行った後、対応するタスク割り当てポリシーに従って前記処理対象タスクが実行されるときの実際の性能を取得し、対応するタスク割り当てポリシー及び取得された実際の性能を訓練データ集合に記憶することを更に含み、前記訓練データ集合は、前記第1ネットワークを更新するために使用される。

【0031】

第3態様において、本開示の実施例は更に、モノのインターネット機器ベースのネットワーク訓練装置を提供し、前記装置は、第1決定ユニットと訓練ユニットとを備え、ここで、

前記第1決定ユニットは、訓練データ集合を決定するように構成され、前記訓練データ集合は、少なくとも1つのタスク割り当てポリシー及び対応する実際の性能を含み、1つの実際の性能は、対応するタスク割り当てポリシーに基づいて実際に実行することによって得られ、

前記訓練ユニットは、前記訓練データ集合に基づいて第1ネットワークを訓練するように構成され、前記第1ネットワークは、タスク割り当てポリシーの性能を予測するために使用される。

【0032】

本開示のいくつかの任意の実施例において、前記装置は更に、処理対象タスクに対応する計算グラフ、及び1つ又は複数のモノのインターネット機器に対応するリソースグラフを決定し、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成するように構成される、第1生成ユニットを備える。

【0033】

本開示のいくつかの任意の実施例において、前記第1生成ユニットは、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成するように構成され、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表し、前記タスク割り当てポリシーは、対応する

10

20

30

40

50

リソースグラフの少なくとも1つのノードを前記計算グラフの各ノードに割り当てるために使用され、リソースサブグラフ内の1つのノードは、1つ又は複数のモノのインターネット機器の能力及びノ又はリソースの少なくとも一部を表し、リソースサブグラフ内の隣接する2つのノードのエッジは、1つ又は複数のモノのインターネット機器の能力及びノ又はリソースの少なくとも一部の間を表す。

【0034】

本開示のいくつかの任意の実施例において、前記第1生成ユニットは、前記計算グラフ内の第1ノードを決定し、前記第1ノードは、リソース需要が最も大きいノードであり、前記リソースグラフ内の少なくとも1つの第2ノードを決定し、前記少なくとも1つの第2ノードは、前記第1ノードのリソース需要を満たすノードであり、各第2ノードに基づいて、1つのリソースサブグラフを決定し、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表すように構成される。

10

【0035】

本開示のいくつかの任意の実施例において、前記訓練ユニットは、少なくとも1つのタスク割り当てポリシーの予測性能及び実際の性能に基づいて、前記第1ネットワークを訓練するように構成される。

【0036】

本開示のいくつかの任意の実施例において、前記訓練ユニットは更に、計算グラフ及び各リソースサブグラフに基づいて、第1ネットワークを用いて各リソースサブグラフに対応する予測性能を得るように構成される。

20

【0037】

本開示のいくつかの任意の実施例において、前記訓練ユニットは、前記第1ネットワークの特徴抽出モジュールを用いて前記計算グラフの特徴を抽出し、第1特徴集合を得、前記特徴抽出モジュールを用いて、前記少なくとも1つのリソースサブグラフの特徴をそれぞれ抽出し、少なくとも1つの第2特徴集合を得、前記第1特徴集合、各第2特徴集合及び前記第1ネットワークの予測モジュールに基づいて、各リソースサブグラフに対応する予測性能を得るように構成される。

【0038】

本開示のいくつかの任意の実施例において、前記訓練ユニットは、前記第1特徴集合及び各第2特徴集合に基づいて、少なくとも1つの第3特徴集合を得、各第3特徴集合は、前記第1特徴集合及び各第2特徴集合を含み、各第3特徴集合及び前記予測モジュールに基づいて、各リソースサブグラフに対応する予測データを得、各リソースサブグラフに対応する予測データに基づいて、各リソースサブグラフに対応する予測性能を得るように構成される。

30

【0039】

本開示のいくつかの任意の実施例において、前記予測データは、
前記処理対象タスクを実行するための予測実行時間長と、
前記処理対象タスクを実行するための予測エネルギー消費量と、
前記処理対象タスクを実行するための予測信頼性と、
のうちの少なくとも1つを含む。

40

【0040】

本開示のいくつかの任意の実施例において、前記訓練ユニットは、所定重みに従って、各リソースサブグラフに対応する予測データに対して重み付け処理を行い、各リソースサブグラフに対応する予測性能を得るように構成される。

【0041】

本開示のいくつかの任意の実施例において、前記訓練ユニットは、各タスク割り当てポリシーの予測性能及び実際の性能に基づいて、前記特徴抽出モジュールと前記予測モジュールを訓練するように構成される。

【0042】

本開示のいくつかの任意の実施例において、前記訓練ユニットは、各タスク割り当てポ

50

リシーの予測性能と実際の性能の間の誤差を逆伝播し、予測性能と実際の性能の間の誤差が所定条件を満たすまで、勾配降下アルゴリズムを利用して第1ネットワークの特徴抽出モジュールと予測モジュールのネットワークパラメータを更新するように構成される。

【0043】

本開示のいくつかの任意の実施例において、前記装置は更に、前記訓練データ集合を更新するように構成される更新ユニットを備え、更新された前記訓練データ集合は、前記第1ネットワークを更新するために使用される。

【0044】

本開示のいくつかの任意の実施例において、前記更新ユニットは、

計算グラフ及びリソースグラフに基づいて、ヒューリスティック法、グラフ探索法、グラフ最適化方法及びサブグラフマッチング方法のうち少なくとも1つの方法を用いて、少なくとも1つのリソースサブグラフを生成し、各リソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、各リソースサブグラフに対応する実際の性能を得、前記計算グラフ、各リソースサブグラフ及び対応する実際の性能を前記訓練データ集合に追加する方式、

10

計算グラフ及びリソースグラフに基づいて、ヒューリスティック法、グラフ探索法、グラフ最適化方法及びサブグラフマッチング方法のうち少なくとも1つの方法を用いて、少なくとも1つのリソースサブグラフを得、第1ネットワークを介して各リソースサブグラフに対応する予測性能を得、前記少なくとも1つのリソースサブグラフから予測性能が最も良いリソースサブグラフを選択し、前記予測性能が最も良いリソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、実際の性能を得、前記計算グラフ、前記予測性能が最も良いリソースサブグラフ及び対応する実際の性能を前記訓練データ集合に追加する方式、

20

計算グラフ及びリソースグラフに基づいて、ランダムウォーク方法を用いて、少なくとも1つのリソースサブグラフを生成し、各リソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、実際の性能を得、前記計算グラフ、少なくとも1つのリソースサブグラフ及び実際の性能を前記訓練データ集合に追加する方式、の少なくとも1つの方式を用いて、前記訓練データ集合を更新するように構成される。

【0045】

第4態様において、本開示の実施例は更に、モノのインターネット機器ベースのタスク割り当て装置を提供し、前記装置は、第2決定ユニット、第2生成ユニット、予測ユニット及びタスク割り当てユニットを備え、ここで、

30

前記第2決定ユニットは、処理対象タスクに対応する計算グラフ、及び1つ又は複数のモノのインターネット機器に対応するリソースグラフを決定するように構成され、

前記第2生成ユニットは、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成するように構成され、

前記予測ユニットは、前記少なくとも1つのタスク割り当てポリシーを第1ネットワークに入力し、各タスク割り当てポリシーに対応する予測性能を取得するように構成され、

前記タスク割り当てユニットは、予測性能が最も良いタスク割り当てポリシーを決定し、決定されたタスク割り当てポリシーに基づいてタスク割り当てを行うように構成される。

40

【0046】

本開示のいくつかの任意の実施例において、前記第2生成ユニットは、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成するように構成され、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表し、前記タスク割り当てポリシーは、対応するリソースグラフの少なくとも1つのノードを前記計算グラフの各ノードに割り当てるために使用され、リソースサブグラフ内の1つのノードは、1つ又は複数のモノのインターネット機器の能力及びノ又はリソースの少なくとも一部を表し、リソースサブグラフ内の隣接する2つのノードのエッジは、1つ又は複数のモノのインターネット機器の能力及びノ又はリソースの少なくとも一部の間の関係を表す。

50

【0047】

本開示のいくつかの任意の実施例において、前記第2生成ユニットは、前記計算グラフ内の第1ノードを決定し、前記第1ノードは、リソース需要が最も大きいノードであり、前記リソースグラフ内の少なくとも1つの第2ノードを決定し、前記少なくとも1つの第2ノードは、前記第1ノードのリソース需要を満たすノードであり、各第2ノードに基づいて、1つのリソースサブグラフを決定し、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表すように構成される。

【0048】

本開示のいくつかの任意の実施例において、前記第1ネットワークは、本開示の実施例の第3態様に記載の装置によって最適化される。

10

【0049】

本開示のいくつかの任意の実施例において、前記予測ユニットは、計算グラフ及び各リソースサブグラフに基づいて、第1ネットワークを用いて各リソースサブグラフに対応する予測性能を得るように構成される。

【0050】

本開示のいくつかの任意の実施例において、前記予測ユニットは、前記第1ネットワークの特徴抽出モジュールを用いて前記計算グラフの特徴を抽出し、第1特徴集合を得、前記特徴抽出モジュールを用いて、前記少なくとも1つのリソースサブグラフの特徴をそれぞれ抽出し、少なくとも1つの第2特徴集合を得、前記第1特徴集合、各第2特徴集合及び前記第1ネットワークの予測モジュールに基づいて、各リソースサブグラフに対応する予測性能を得るように構成される。

20

【0051】

本開示のいくつかの任意の実施例において、前記予測ユニットは、前記第1特徴集合及び各第2特徴集合に基づいて、少なくとも1つの第3特徴集合を得、各第3特徴集合は、前記第1特徴集合及び各第2特徴集合を含み、各第3特徴集合及び前記予測モジュールに基づいて、各リソースサブグラフに対応する予測データを得、各リソースサブグラフに対応する予測データに基づいて、各リソースサブグラフに対応する予測性能を得るように構成される。

【0052】

本開示のいくつかの任意の実施例において、前記予測データは、
前記処理対象タスクを実行するための予測実行時間長と、
前記処理対象タスクを実行するための予測エネルギー消費量と、
前記処理対象タスクを実行するための予測信頼性と、
のうちの少なくとも1つを含む。

30

【0053】

本開示のいくつかの任意の実施例において、前記予測ユニットは、所定重みに従って、各リソースサブグラフに対応する予測データに対して重み付け処理を行い、各リソースサブグラフに対応する予測性能を得るように構成される。

【0054】

本開示のいくつかの任意の実施例において、前記装置は更に、タスク割り当てを行った後、対応するタスク割り当てポリシーに従って前記処理対象タスクが実行されるときの実際の性能を取得し、対応するタスク割り当てポリシー及び取得された実際の性能を訓練データ集合に記憶するように構成される取得ユニットを備え、前記訓練データ集合は、前記第1ネットワークを更新するために使用される。

40

【0055】

第5態様において、本開示の実施例は更に、プロセッサによって実行されるとき、前記プロセッサに本開示の実施例の上記の第1態様又は第2態様に記載の方法を実現させるためのコンピュータプログラムが記憶された、コンピュータ可読記憶媒体を提供する。

【0056】

第6態様において、本開示の実施例は更に、メモリと、プロセッサと、メモリに記憶さ

50

れプロセッサで実行可能なコンピュータプログラムと、を備える電子機器を提供し、前記プロセッサは、前記プログラムを実行するとき、本開示の実施例の上記の第1態様又は第2態様に記載の方法を実現する。

【0057】

本開示の実施例の技術的解決策において、一方では、訓練データ集合を決定することにより、前記訓練データ集合に基づいて第1ネットワークを訓練し、前記訓練データ集合は、少なくとも1つのタスク割り当てポリシー及び対応する実際の性能を含み、1つの実際の性能は、対応するタスク割り当てポリシーに基づいて実際に実行することによって得られ、前記実際の性能は、性能の実際の値であり、前記第1ネットワークは、タスク割り当てポリシーの性能を予測するために使用され、他方では、処理対象タスクに対応する計算グラフ、及び1つ又は複数のモノのインターネット機器に対応するリソースグラフを決定し、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成し、前記少なくとも1つのタスク割り当てポリシーを第1ネットワークに入力し、各タスク割り当てポリシーに対応する予測性能を得、前記予測性能は、性能の予測値であり、予測性能が最も良いタスク割り当てポリシーを決定し、決定されたタスク割り当てポリシーに基づいてタスク割り当てを行う。このように、第1ネットワークの訓練により、特定のタスク割り当てポリシーを用いて、処理対象タスクを実行するためのシステム性能の予測を実現し、それにより、最適な割り当てポリシーを決定し、タスク割り当てプロセスにおける処理対象タスクと機器の利用可能なリソース間の最適なマッチングを実現し、それにより、リソースの利用を最大化し、システム性能を向上させることができる。

本発明は、例えば、以下を提供する。

(項目1)

モノのインターネット機器ベースのネットワーク訓練方法であって、

訓練データ集合を決定し、前記訓練データ集合に基づいて第1ネットワークを訓練することを含み、前記訓練データ集合は、少なくとも1つのタスク割り当てポリシー及び対応する実際の性能を含み、1つの実際の性能は、対応するタスク割り当てポリシーに基づいて実際に実行することによって得られ、前記第1ネットワークは、タスク割り当てポリシーの性能を予測するために使用される、ネットワーク訓練方法。

(項目2)

前記ネットワーク訓練方法は、

処理対象タスクに対応する計算グラフ、及び1つ又は複数のモノのインターネット機器に対応するリソースグラフを決定し、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成することを更に含む、

項目1に記載のネットワーク訓練方法。

(項目3)

前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成することは、

前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成することを含み、各リソースサブグラフは、1つのタスク割り当てポリシーを含み、前記タスク割り当てポリシーは、対応するリソースグラフの少なくとも1つのノードを前記計算グラフの各ノードに割り当てるために使用され、リソースサブグラフ内の1つのノードは、1つ又は複数のモノのインターネット機器の能力の少なくとも一部を表し、リソースサブグラフ内の隣接する2つのノードのエッジは、1つ又は複数のモノのインターネット機器の能力の少なくとも一部の間を関係を表す、

項目2に記載のネットワーク訓練方法。

(項目4)

前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成することは、

前記計算グラフ内の第1ノードを決定することであって、前記第1ノードは、リソース

10

20

30

40

50

需要が最も大きいノードである、ことと、

前記リソースグラフ内の少なくとも1つの第2ノードを決定することであって、前記少なくとも1つの第2ノードは、前記第1ノードのリソース需要を満たすノードである、ことと、

各第2ノードに基づいて、1つのリソースサブグラフを決定することであって、各リソースサブグラフは1つのタスク割り当てポリシーを含む、ことと、を含む、

項目3に記載のネットワーク訓練方法。

(項目5)

前記第1ネットワークを訓練することは、

少なくとも1つのタスク割り当てポリシーの予測性能及び実際の性能に基づいて、前記第1ネットワークを訓練することを含み、

計算グラフ及び各リソースサブグラフに基づいて、第1ネットワークを用いて各リソースサブグラフに対応する予測性能を得る、

項目1に記載のネットワーク訓練方法。

(項目6)

前記第1ネットワークを用いて各リソースサブグラフに対応する予測性能を得ることは、

前記第1ネットワークの特徴抽出モジュールを用いて前記計算グラフの特徴を抽出し、第1特徴集合を得ることと、

前記特徴抽出モジュールを用いて、前記少なくとも1つのリソースサブグラフの特徴をそれぞれ抽出し、少なくとも1つの第2特徴集合を得ることと、

前記第1特徴集合、各第2特徴集合及び前記第1ネットワークの予測モジュールに基づいて、各リソースサブグラフに対応する予測性能を得ることと、を含む、

項目5に記載のネットワーク訓練方法。

(項目7)

前記第1特徴集合、各第2特徴集合及び前記第1ネットワークの予測モジュールに基づいて、各リソースサブグラフに対応する予測性能を得ることは、

前記第1特徴集合及び各第2特徴集合に基づいて、少なくとも1つの第3特徴集合を得ることであって、各第3特徴集合は、前記第1特徴集合及び各第2特徴集合を含む、ことと、

各第3特徴集合及び前記予測モジュールに基づいて、各リソースサブグラフに対応する予測データを得、各リソースサブグラフに対応する予測データに基づいて、各リソースサブグラフに対応する予測性能を得ることと、を含む、

項目6に記載のネットワーク訓練方法。

(項目8)

前記第1ネットワークを訓練することは、

各タスク割り当てポリシーの予測性能及び実際の性能に基づいて、前記特徴抽出モジュールと前記予測モジュールを訓練することを含む、

項目7に記載のネットワーク訓練方法。

(項目9)

前記特徴抽出モジュールと前記予測モジュールを訓練することは、

各タスク割り当てポリシーの予測性能と実際の性能の間の誤差を逆伝播し、予測性能と実際の性能の間の誤差が所定条件を満たすまで、勾配降下アルゴリズムを利用して第1ネットワークの特徴抽出モジュールと予測モジュールのネットワークパラメータを更新することを含む、

項目8に記載のネットワーク訓練方法。

(項目10)

前記ネットワーク訓練方法は、前記訓練データ集合を更新することを更に含み、

前記訓練データ集合を更新することは、

計算グラフ及びリソースグラフに基づいて、ヒューリスティック法、グラフ探索法、グラフ最適化方法及びサブグラフマッチング方法のうちの少なくとも1つの方法を用いて、

10

20

30

40

50

少なくとも1つのリソースサブグラフを生成し、各リソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、各リソースサブグラフに対応する実際の性能を得、前記計算グラフ、各リソースサブグラフ及び対応する実際の性能を前記訓練データ集合に追加すること、

計算グラフ及びリソースグラフに基づいて、ヒューリスティック法、グラフ探索法、グラフ最適化方法及びサブグラフマッチング方法のうち少なくとも1つの方法を用いて、少なくとも1つのリソースサブグラフを生成し、第1ネットワークを介して各リソースサブグラフに対応する予測性能を得、前記少なくとも1つのリソースサブグラフから予測性能が最も良いリソースサブグラフを選択し、前記予測性能が最も良いリソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、実際の性能を得、前記計算グラフ、前記予測性能が最も良いリソースサブグラフ及び対応する実際の性能を前記訓練データ集合に追加すること、

10

計算グラフ及びリソースグラフに基づいて、ランダムウォーク方法を用いて、少なくとも1つのリソースサブグラフを生成し、各リソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、実際の性能を得、前記計算グラフ、少なくとも1つのリソースサブグラフ及び実際の性能を前記訓練データ集合に追加すること、のうちの少なくとも1つを含む、

項目1に記載のネットワーク訓練方法。

(項目11)

モノのインターネット機器ベースのタスク割り当て方法であって、

20

処理対象タスクに対応する計算グラフ、及び1つ又は複数のモノのインターネット機器に対応するリソースグラフを決定することと、

前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成することと、

前記少なくとも1つのタスク割り当てポリシーを第1ネットワークに入力し、各タスク割り当てポリシーに対応する予測性能を取得することと、

予測性能が最も良いタスク割り当てポリシーを決定し、決定されたタスク割り当てポリシーに基づいてタスク割り当てを行うことと、を含む、タスク割り当て方法。

(項目12)

前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成することは、

30

前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成することを含み、各リソースサブグラフは、1つのタスク割り当てポリシーを含み、前記タスク割り当てポリシーは、対応するリソースグラフの少なくとも1つのノードを前記計算グラフの各ノードに割り当てるために使用され、リソースサブグラフ内の1つのノードは、1つ又は複数のモノのインターネット機器の能力の少なくとも一部を表し、リソースサブグラフ内の隣接する2つのノードのエッジは、1つ又は複数のモノのインターネット機器の能力の少なくとも一部の間の関係を表す、

項目11に記載のタスク割り当て方法。

(項目13)

40

前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成することは、

前記計算グラフ内の第1ノードを決定することであって、前記第1ノードは、リソース需要が最も大きいノードである、ことと、

前記リソースグラフ内の少なくとも1つの第2ノードを決定することであって、前記少なくとも1つの第2ノードは、前記第1ノードのリソース需要を満たすノードである、ことと、

各第2ノードに基づいて、1つのリソースサブグラフを決定することであって、各リソースサブグラフは1つのタスク割り当てポリシーを含む、ことと、を含む、

項目12に記載のタスク割り当て方法。

50

(項目14)

前記各タスク割り当てポリシーに対応する予測性能を取得することは、
計算グラフ及び各リソースサブグラフに基づいて、第1ネットワークを用いて各リソースサブグラフに対応する予測性能を得ることを含む、
項目11に記載のタスク割り当て方法。

(項目15)

前記第1ネットワークを用いて各リソースサブグラフに対応する予測性能を得ることは、
前記第1ネットワークの特徴抽出モジュールを用いて前記計算グラフの特徴を抽出し、
第1特徴集合を得ることと、
前記特徴抽出モジュールを用いて、前記少なくとも1つのリソースサブグラフの特徴を
それぞれ抽出し、少なくとも1つの第2特徴集合を得ることと、
前記第1特徴集合、各第2特徴集合及び前記第1ネットワークの予測モジュールに基づ
いて、各リソースサブグラフに対応する予測性能を得ることと、を含む、
項目14に記載のタスク割り当て方法。

10

(項目16)

前記第1特徴集合、各第2特徴集合及び前記第1ネットワークの予測モジュールに基づ
いて、各リソースサブグラフに対応する予測性能を得ることは、
前記第1特徴集合及び各第2特徴集合に基づいて、少なくとも1つの第3特徴集合を得
ることとあって、各第3特徴集合は、前記第1特徴集合及び各第2特徴集合を含む、こと
と、
各第3特徴集合及び前記予測モジュールに基づいて、各リソースサブグラフに対応する
予測データを得、各リソースサブグラフに対応する予測データに基づいて、各リソースサ
ブグラフに対応する予測性能を得ることと、を含む、
項目15に記載のタスク割り当て方法。

20

(項目17)

前記タスク割り当て方法は、
タスク割り当てを行った後、対応するタスク割り当てポリシーに従って前記処理対象タ
スクが実行されるときの実際の性能を取得し、対応するタスク割り当てポリシー及び取得
された実際の性能を訓練データ集合に記憶することを更に含み、前記訓練データ集合は、
前記第1ネットワークを更新するために使用される、
項目11に記載のタスク割り当て方法。

30

(項目18)

モノのインターネット機器ベースのネットワーク訓練装置であって、第1決定ユニット
と訓練ユニットとを備え、
前記第1決定ユニットは、訓練データ集合を決定するように構成され、前記訓練データ
集合は、少なくとも1つのタスク割り当てポリシー及び対応する実際の性能を含み、1つ
の実際の性能は、対応するタスク割り当てポリシーに基づいて実際に実行することによっ
て得られ、
前記訓練ユニットは、前記訓練データ集合に基づいて第1ネットワークを訓練するよう
に構成され、前記第1ネットワークは、タスク割り当てポリシーの性能を予測するために
使用される、ネットワーク訓練装置。

40

(項目19)

前記ネットワーク訓練装置は更に、処理対象タスクに対応する計算グラフ、及び1つ又
は複数のモノのインターネット機器に対応するリソースグラフを決定し、前記計算グラフ
及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成す
るように構成される第1生成ユニットを備える、
項目18に記載のネットワーク訓練装置。

(項目20)

前記第1生成ユニットは、前記計算グラフ及び前記リソースグラフに基づいて、少なく
とも1つのリソースサブグラフを生成するように構成され、各リソースサブグラフは、1

50

つのタスク割り当てポリシーを含み、前記タスク割り当てポリシーは、対応するリソースグラフの少なくとも1つのノードを前記計算グラフの各ノードに割り当てるために使用され、リソースサブグラフ内の1つのノードは、1つ又は複数のモノのインターネット機器の能力の少なくとも一部を表し、リソースサブグラフ内の隣接する2つのノードのエッジは、1つ又は複数のモノのインターネット機器の能力の少なくとも一部の間を表す、
項目19に記載のネットワーク訓練装置。

(項目21)

前記第1生成ユニットは、前記計算グラフ内の第1ノードを決定し、前記第1ノードは、リソース需要が最も大きいノードであり、前記リソースグラフ内の少なくとも1つの第2ノードを決定し、前記少なくとも1つの第2ノードは、前記第1ノードのリソース需要を満たすノードであり、各第2ノードに基づいて、1つのリソースサブグラフを決定し、各リソースサブグラフは、1つのタスク割り当てポリシーを含むように構成される、
項目19に記載のネットワーク訓練装置。

(項目22)

モノのインターネット機器ベースのタスク割り当て装置であって、第2決定ユニット、第2生成ユニット、予測ユニット及びタスク割り当てユニットを備え、

前記第2決定ユニットは、処理対象タスクに対応する計算グラフ、及び1つ又は複数のモノのインターネット機器に対応するリソースグラフを決定するように構成され、

前記第2生成ユニットは、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成するように構成され、

前記予測ユニットは、前記少なくとも1つのタスク割り当てポリシーを第1ネットワークに入力し、各タスク割り当てポリシーに対応する予測性能を取得するように構成され、

前記タスク割り当てユニットは、予測性能が最も良いタスク割り当てポリシーを決定し、決定されたタスク割り当てポリシーに基づいてタスク割り当てを行うように構成される、
タスク割り当て装置。

(項目23)

プロセッサによって実行されるとき、前記プロセッサに項目1ないし10のいずれか一項に記載のモノのインターネット機器ベースのネットワーク訓練方法を実現させ、又は、プロセッサによって実行されるとき、前記プロセッサに項目11ないし17のいずれか一項に記載のモノのインターネット機器ベースのタスク割り当て方法を実現させるためのコンピュータプログラムが記憶された、コンピュータ可読記憶媒体。

(項目24)

メモリと、プロセッサと、メモリに記憶されプロセッサで実行可能なコンピュータプログラムと、を備える電子機器であって、

前記プロセッサは、前記プログラムを実行して、項目1ないし10のいずれか一項に記載のモノのインターネット機器ベースのネットワーク訓練方法を実現し、又は、前記プログラムは、前記プロセッサによって実行されるとき、項目11ないし17のいずれか一項に記載のモノのインターネット機器ベースのタスク割り当て方法を実現する、電子機器。

【図面の簡単な説明】

【0058】

【図1】本開示の実施例に適用される、選択的なシーンの概略図である。

【0059】

【図2】本開示の実施例のモノのインターネット機器ベースのネットワーク訓練方法の例示的なフローチャートである。

【0060】

【図3】本開示の実施例における、リソースサブグラフの生成方式の例示的なフローチャートである。

【0061】

【図4】本開示の実施例における、リソースサブグラフの生成の概略図である。

【0062】

10

20

30

40

50

【図5】本開示の実施例における、予測性能の取得方式の例示的なフローチャートである。

【0063】

【図6a】本開示の実施例における、計算グラフの特徴抽出の概略図である。

【0064】

【図6b】本開示の実施例における、リソースサブグラフの特徴抽出の概略図である。

【0065】

【図7】本開示の実施例のモノのインターネット機器ベースのタスク割り当て方法の例示的なフローチャートである。

【0066】

【図8】本開示の実施例のモノのインターネット機器ベースのタスク割り当て方法の概略図である。

10

【0067】

【図9】本開示の実施例のタスク割り当てシステムの構成の概略図である。

【0068】

【図10】本開示の実施例モノのインターネット機器ベースのネットワーク訓練装置の構成の例示的な構造図1である。

【0069】

【図11】本開示の実施例モノのインターネット機器ベースのネットワーク訓練装置の構成の例示的な構造図2である。

【0070】

20

【図12】本開示の実施例モノのインターネット機器ベースのネットワーク訓練装置の構成の例示的な構造図3である。

【0071】

【図13】本開示の実施例モノのインターネット機器ベースのタスク割り当て装置の構成の例示的な構造図である。

【0072】

【図14】本開示の実施例の電子機器のハードウェア構成の例示的な構造図である。

【発明を実施するための形態】

【0073】

以下では、図面及び特定の実施例を参照して、本開示について更に詳細に説明する。

30

【0074】

一方では、深層学習技術の画期的な進歩、及び5G技術の推進・普及に伴い、車両ネットワークキング、スマート高齢者介護、スマートコミュニティ、産業用インターネットなどの分野で、ますます多くのインテリジェントなサービスが提供されることになり、これらのインテリジェントなサービスの実現は、人工知能技術と深層学習モデルに依存することが多い。一方では、1つ又は複数のモノのインターネット機器の数とインテリジェンス程度の急速な増加に伴い、リソースに制約があり、高度の異機種であるモノのインターネット機器を最大限に活用するために、広く分散したモノのインターネット機器上の遊休リソース（フリーリソース、アイドルリソース、利用可能リソース、遊休能力、フリー能力、利用可能能力、アイドル能力とも呼ばれる）を利用し、リソース共有と機器連携を通じて、分散方式で計算集約的な計算タスクを実行することを考えられることができる。これを鑑みて、本開示の実施例は、主に、異機種モノのインターネット機器間のリソーススケジューリング、エンドツーエンドの訓練可能なネットワークモデル、自動的に最適化を実現し、高性能かつインテリジェントに自己適応するタスク割り当て方法を構築することを目的とする。長期的に最適化されたリソース管理及びタスクスケジューリングポリシーを学習することにより、最適なシステム性能を得ることができるタスク割り当て方式で、計算グラフ内のノードに適切な計算能力、ストレージ及び通信リソースを割り当て、機器間連携の分散型機械学習（例えば、深層モデルの訓練や推論）の実現を促進し、モノのインターネットシーンにおけるインテリジェントアプリケーションとサービスの実現に更に貢献する。

40

50

【0075】

図1は、本開示の実施例に適用される、選択的なシーンの概略図であり、図1に示すように、スマートホームサービス(Smart Home Service)は、ホームサービスロボット(Home Service Robot)、インテリジェントモニタリング(Intelligent monitoring)、仮想現実(VR:Virtual Reality)、インテリジェント制御(Intelligent control)などのサービスを含み得るが、これらに限定されない。モノのインターネット機器を用いて、ビデオ、画像、音声、テキストなどのデータを含む構造化及び非構造化データを収集することができ、収集されたデータを設計されたネットワークモデルに入力し、モノのインターネット機器上のハードウェアリソースを利用して、対応する計算タスクを実行することにより、AIアプリケーション(AI application)中の行動識別(Action Recognition)、自然言語処理(Natural Language Processing)、画像処理(Image Recognition)、顔識別(Face Recognition)などの様々なインテリジェント機能を実現することができる。ここで、深層ニューラルネットワーク(DNN:Deep Neural Networks)、畳み込みニューラルネットワーク(CNN:Convolutional Neural Networks)、リカレントニューラルネットワーク(RNN:Recurrent Neural Network)、長期短期記憶(LSTM:Long Short-Term Memory)ネットワーク、グラフ畳み込みネットワーク(GCN:Graph Convolutional Networks)、グラフニューラルネットワーク(GNN:Graph Neural Networks)などの計算タスクにより上記の機能を実現することができる。計算タスクを実現するための上記のネットワークの様々な抽象的な演算を分割して、畳み込み(Conv)やプーリング(Pooling)などの一連の演算子を得、これらは、処理対象タスク内の特定の種類の演算を表し、各演算子は、演算子ライブラリ(Operator library)を形成することができる。

10

20

【0076】

シーンは更に、複数のエッジ側のモノのインターネット機器(IoT Device)を含み、各モノのインターネット機器は、異なる能力を有することができる。前記能力は、計算リソース(Computation Resource)、ストレージリソース(Storage/Memory Resource)、通信リソース(Communication Resource)によって具現でき、上記の計算リソースは、利用可能又は遊休の計算リソースを指してもよく、計算リソースは、例えば、中央処理装置(CPU:Central Processing Unit)リソース、グラフィックプロセッサ(GPU:Graphics Processing Unit)リソース、フィールドプログラマブルゲートアレイ(FPGA:Field-Programmable Gate Array)リソース、デジタル信号プロセッサ(DSP:Digital Signal Processor)リソースなどを含み得、上記のストレージリソースは、利用可能又は遊休のストレージリソースを指してもよく、ストレージリソースは例えば、メモリ(Memorizer)リソース、キャッシュメモリ(Cache)リソース、ランダムアクセスメモリ(RAM:Random Access Memory)などを含み得、上記の通信リソースは、利用可能又は遊休の通信リソースを指してもよい。そして、異なるモノのインターネット機器は、異なる計算リソース、ストレージリソース、通信リソースのうち少なくとも1つのリソースを提供することができ、各モノのインターネット機器の計算リソース、ストレージリソース、通信リソースは、形成リソースプールを形成することができる。

30

40

【0077】

本実施例において、計算タスクの各演算子に対応するノードに抽象化することにより、計算グラフを生成し、各モノのインターネット機器の能力に対応するノードに抽象化することにより、リソースグラフを生成し、計算グラフ及びリソースグラフに基づいて、リソ

50

ースサブグラフの構築 (Resource Sub-graph Construction) を行い、計算グラフ及びリソースサブグラフに対して特徴抽出 (Feature Extraction) を行い、抽出された特徴に基づいて、その中に暗黙的に含まれるタスク割り当てポリシーに対して性能予測 (Performance Prediction) などの方式を行い、それにより、タスクのインテリジェントな割り当てを実現する。

【0078】

なお、上記の図1に示すスマートホームサービスシーンは、本開示の実施例の技術的解決策が適用される任意の適用シーンに過ぎず、他の適用シーンも、本開示の実施例の保護範囲内にあり得、本開示の実施例は、特に限定されない。

【0079】

本開示の実施例は、1つ又は複数のモノのインターネット機器ベースのネットワーク訓練方法を提供し、様々な電子機器に適用され、前記電子機器は、固定機器及び/又はモバイル機器を含むが、これらに限定されない。例えば、前記固定機器は、パーソナルコンピュータ (PC: Personal Computer) 又はサーバなどを含むが、これらに限定されなく、前記サーバは、クラウドサーバ又は一般的なサーバであってもよい。前記モバイル機器は、携帯電話、タブレットコンピュータ又はウェアラブル機器のうちの1つ又は複数を含むが、これらに限定されない。

【0080】

図2は、本開示の実施例のモノのインターネット機器ベースのネットワーク訓練方法の例示的なフローチャートであり、図2に示すように、前記方法は、以下のステップを含む。

【0081】

ステップ101において、訓練データ集合を決定し、前記訓練データ集合は、少なくとも1つのタスク割り当てポリシー及び対応する実際の性能を含み、1つの実際の性能は、対応するタスク割り当てポリシーに基づいて実際に実行することによって得られる。

【0082】

ステップ102において、前記訓練データ集合に基づいて第1ネットワークを訓練し、前記第1ネットワークは、タスク割り当てポリシーの性能を予測するために使用される。

【0083】

本実施形態において、前記タスク割り当てポリシーは、処理対象タスクを少なくとも1つのモノのインターネット機器によって実行されるポリシーに割り当ててことを示し、言い換えれば、前記タスク割り当てポリシーを用いて、少なくとも1つのモノのインターネット機器を決定し、少なくとも1つのモノのインターネット機器を用いて、タスク割り当てポリシーの指示に従って処理対象タスクを実行することができる。任意選択的に、前記タスク割り当てポリシーは、タスク割り当て方法、タスク割り当て方式、タスクスケジューリングポリシー、タスクスケジューリング方法、タスクスケジューリング方式、タスク割り付けポリシー、タスク割り付け方法、タスク割り付け方式などのうちの1つとも呼ばれる。

【0084】

本実施形態において、異なるタスク割り当てポリシーを用いて処理対象タスクを実際に実行することにより、同じ又は異なるシステム性能、即ち、上記の実際の性能を得ることができる。前記実際の性能は、処理対象タスクを実際に実行する際のシステムの性能 (例えば、実行時間長、エネルギー消費量、信頼性など) を示す。本実施例は、訓練データ集合内の各タスク割り当てポリシー及び対応する実際の性能を用いて第1ネットワークを訓練する。

【0085】

本実施形態において、システムにおける各モノのインターネット機器は、異機種モノのインターネット機器であってもよい。ここで、異機種モノのインターネット機器とは、複数のモノのインターネット機器とサーバを含む1つのネットワークにおいて、1つのモノのインターネット機器のハードウェアが、他のモノのインターネット機器のハードウェアと異なること、及び/又は、1つのモノのインターネット機器のサーバが、他のモノのイ

10

20

30

40

50

インターネット機器のサーバと異なることを指す。ここで、1つのモノのインターネット機器のハードウェアが他のモノのインターネット機器のハードウェアと異なることは、1つのモノのインターネット機器と他のモノのインターネット機器の計算リソース、ストレージリソースの少なくとも1つのリソースに対応するハードウェアのモデル又は種類が異なることを指す。例えば、計算リソースに対応するハードウェアを例にとると、1つのモノのインターネット機器のCPU、GPU、バスインターフェースチップ(BIC: Bus Interface Chip)、DSP、FPGA、特定用途向け集積回路(ASIC: Application Specific Integrated Circuit)、テンソル処理ユニット(TPU: Tensor Processing Unit)、人工知能(AI: Artificial Intelligence)チップのうちの少なくとも1つのハードウェアのモデルが、他のモノのインターネット機器の対応するハードウェアのモデルと異なり、更に例えば、ストレージリソースに対応するハードウェアを例にとると、1つのモノのインターネット機器のRAM、読み取り専用メモリ(ROM: Read Only Memory)、Cacheのうちの少なくとも1つのハードウェアのモデルが、他のモノのインターネット機器の対応するハードウェアのモデルと異なる。1つのモノのインターネット機器のサーバが、他のモノのインターネット機器のサーバと異なることは、1つのモノのインターネット機器に対応するバックエンドプログラムが、他のモノのインターネット機器に対応するバックエンドプログラムと異なることを指し、ここで、前記バックエンドプログラムは、オペレーティングシステムを含んでもよく、つまり、1つのモノのインターネット機器に対応するオペレーティングシステムは、他のモノのインターネット機器に対応するオペレーティングシステムと異なり、言い換えれば、2つのモノのインターネット機器の間にはソフトウェアレベルで違いがある。

【0086】

例示的に、前記モノのインターネット機器は、携帯電話、PC、ウェアラブルスマート機器、スマートゲートウェイ、コンピューティングボックスなどを含み得、前記PCは、デスクトップコンピュータ、ノートブックコンピュータ、タブレットコンピュータなどを含み得、前記ウェアラブルスマート機器は、スマートウォッチ、スマートメガネなどを含み得る。

【0087】

本開示のいくつかの任意の実施例において、前記方法は、処理対象タスクに対応する計算グラフ、及び1つ又は複数のモノのインターネット機器に対応するリソースグラフを決定し、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成することを更に含む。

【0088】

本実施例において、計算グラフは、処理対象タスクを含み得、計算グラフは、少なくとも1つのノードを含み得、各ノードは、処理対象タスク内の特定の演算又は演算子に対応し、ノード間のエッジは、隣接するノード間の関係を表す。例示的に、計算グラフは、3つのノードを含み、3つのノードは順次接続されている場合、処理対象タスクが3つのノードに対応する演算子によって実現されることを示すことができ、最初のノードが対応する最初の演算子を通して処理された後、出力データは2番目のノードに送信され、更に、2番目のノードに対応する2番目の演算子を通して処理され、処理されたデータは3番目のノードに送信され、3番目のノードに対応する3番目の演算子を通して処理され、それにより、処理対象タスクを実現する。

【0089】

本実施形態において、リソースグラフは、システムにおける各モノのインターネット機器の能力(リソースとも呼ばれる)を含み得る。モノのインターネット機器の能力は少なくとも、計算能力、記憶能力及び通信能力のうちの少なくとも1つを含む。例示的に、リソースグラフは、少なくとも1つのノードを含み得、各ノードは、1つ又は複数のモノのインターネット機器の能力及び/又はリソースの少なくとも一部に対応し、一例において、1つのノードは、1つのモノのインターネット機器のすべての能力(例えば、計算能力

10

20

30

40

50

、記憶能力及び通信能力など)を表すことができ、他の例において、1つのノードは、1つのモノのインターネット機器の能力の一部を表すことができ、例えば、1つのモノのインターネット機器の計算能力又は記憶能力のみを表し、又は、1つのモノのインターネット機器の計算能力の一部及び/又は記憶能力の一部のみを表す。ノード間のエッジは、1つ又は複数のモノのインターネット機器の能力及び/又はリソースの少なくとも一部間の関係を示す。

【0090】

いくつかの選択的な実施形態において、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成することは、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成することを含み、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表し、前記タスク割り当てポリシーは、対応するリソースグラフの少なくとも1つのノードを前記計算グラフの各ノードに割り当てるために使用され、リソースサブグラフ内の1つのノードは、1つ又は複数のモノのインターネット機器の能力及び/又はリソースの少なくとも一部を表し、リソースサブグラフ内の隣接する2つのノードのエッジは、1つ又は複数のモノのインターネット機器の能力及び/又はリソースの少なくとも一部間の関係を表す。

10

【0091】

本実施例において、1つのタスク割り当てポリシーは、計算グラフ内の各ノードに、対応するリソースグラフの少なくとも1つのノードを割り当てるために使用され、又は処理対象タスクを少なくとも1つのモノのインターネット機器に割り当て又はマッピングするか、又は、処理対象タスクとモノのインターネット機器との間マッチングのため、又は、処理対象タスクとリソースとの間のマッチングのために使用される。

20

【0092】

実際の適用では、前記計算グラフの各ノードに割り当てられた前記リソースグラフの少なくとも1つのノードは、同じであってもよいし異なってもよく、つまり、1つのモノのインターネット機器は、自身の能力及び/又はリソースの少なくとも一部を用いて複数の演算子に対応する計算ユニットを実現することができ、同時に、複数のモノのインターネット機器は、連携して1つの演算子に対応する計算ユニットを実現することができる。更に、計算依存関係(即ち、データ依存関係)を有しない計算グラフ内のノード(又は演算子)は、同じ又は異なるモノのインターネット機器で並列に実行(又は演算、計算)されてもよい。

30

【0093】

例示的に、1つのタスク割り当てポリシーは、1つのリソースサブグラフによって具現でき、言い換えれば、前記訓練データ集合は、計算グラフ、少なくとも1つのリソースサブグラフ及び対応する実際の性能を含み得る。本実施例において、処理対象タスクによって構築された計算グラフ及び遊休リソースを含む複数の異機種モノのインターネット機器によって構築されたリソースグラフに基づいて、計算リソース、ストレージリソース及び通信リソースに対する処理対象タスクの需要と各モノのインターネット機器上の利用可能リソース又は能力との関係に基づいて、完全なリソースグラフから、少なくとも1つのリソースサブグラフを生成し、つまり、少なくとも1つのタスク割り当てポリシーを生成する。リソースサブグラフの構築により、モノのインターネット機器上のフリーリソースの最大限の活用及び細かい粒度のタスク割り当てと最適化が実現される。

40

【0094】

任意選択的に、本実施例におけるリソースグラフは、リソース知識グラフ又はリソース知識アトラスとも呼ばれ、リソースサブグラフは、リソース知識サブグラフ又はリソース知識サブアトラスなどとも呼ばれる。

【0095】

任意選択的に、図3は、本開示の実施例における、リソースサブグラフの生成方式の例示的なフローチャートであり、図3に示すように、リソースサブグラフの生成方法は、以

50

下のステップを含み得る。

【0096】

ステップ201において、前記計算グラフ内の第1ノードを決定し、前記第1ノードは、リソース需要が最も大きいノードである。

【0097】

ステップ202において、前記リソースグラフ内の少なくとも1つの第2ノードを決定し、前記少なくとも1つの第2ノードは、前記第1ノードのリソース需要を満たすノードである。

【0098】

ステップ203において、各第2ノードに基づいて、1つのリソースサブグラフを決定し、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表す。

10

【0099】

例示的に、図4は、本開示の実施例における、リソースサブグラフの生成の概略図であり、図4を参照すると、まず、1つの計算グラフ内の各ノードに番号が付けられる。

【0100】

例示的に、計算グラフ内のノードは、統一されたルールに基づいてラベル付けされる。例えば、まず、計算グラフ内のノードの数が最も多いブランチを決定し、当該ブランチ内の各ノードの優先順位順に番号を付ける。例えば、図4におけるノードの数が最も多いブランチに5個のノードがあり、5個のノードを有するブランチには、優先順位順に番号が付けられ、更に、ノードの数が2番目に多いブランチ内のすべてのノードに番号を付け、計算グラフ内のすべてのノードに番号が付けられるまで、この作業を繰り返す。

20

【0101】

次に、計算グラフのすべてのノードのうちの第1ノードを決定し、前記第1ノードは、リソース需要が最も大きいノードであり、前記第1ノードは、ボトルネックノードとも呼ばれる。例えば、図4における計算グラフ内の番号4のノードが最大リソース需要を有し、番号4のノードを第1ノード又はボトルネックノードとして決定する。ここで、リソース需要が最も大きいノードは、計算リソース、ストレージリソース及び通信リソースのうちの少なくとも1つのリソースの需要が最も大きいノードであってもよい。

【0102】

更に、リソースグラフ内の少なくとも1つの第2ノードを決定し、つまり、第1ノード（又はボトルネックノード）のために適切なリソースノード（機器ノード、又は能力ノードとも呼ばれる）を割り当て、処理対象タスクの実行に利用可能なリソースを提供する。例示的に、第1ノード（又はボトルネックノード）のリソース需要を満たすことができるリソースグラフ内のノードを、いずれも第2ノードとして決定し、例えば、図4のリソースグラフ内の3つの番号4のノードが、いずれも第1ノード（又はボトルネックノード）のリソース需要を満たすことができる場合、3つの番号4のノードを第2ノードとして決定する。通常、リソースグラフにおいて、第1ノード（又はボトルネックノード）のリソース需要を満たすことができるリソースノードの数は1つ以上であるため、生成されるリソースサブグラフも1つに限られていない。

30

40

【0103】

また、次に、リソースグラフにおいて、第1ノード（又はボトルネックノード）に対応する各リソースノードを出発点（開始ノード）とし、例えば、図4におけるリソースグラフ上の右側の4に番号付けられたリソースノード（ノードV3と表記する）を出発点として、リソースグラフにおいて、開始ノードに隣接する他のリソースノード（例えば、図4のリソースグラフにおけるノードV1、V4及びV5）を検索することができる。対応する作業負荷のリソース需要を満たすために、計算グラフにおける第1ノードから1ホップ離れたノード（例えば、計算グラフ内の3、5、6に番号付けられたノード）に適切なリソースノードを割り当ててもよい。例えば、リソースグラフ内のノードV1を計算グラフ内のノード3に割り当て、リソースグラフ内のノードV4を計算グラフ内のノード6に

50

割り当て、リソースグラフ内のノードV5を計算グラフ内のノード5に割り当てる。更に、計算グラフにおいて第1ノードから2ホップ離れたノード(例えば、計算グラフ内の2に番号付けられたノード)にリソースノードを割り当てる。例えば、リソースグラフ内のノードV2を計算グラフ内のノード2に割り当てる。計算グラフ内のすべてのノードに、リソースグラフ内のリソースノードが割り当てられるまで、作業を繰り返す。

【0104】

ここで、第1ノード(又はボトルネックノード)のリソース需要を満たすリソースグラフ内の各リソースノードに対して、上記のステップに従ってリソースノードの割り当てを実行し、それにより、図4における右側に示す3つのタスク割り当てポリシーを得ることができ、即ち、3つのリソースサブグラフを得ることができる。複数のタスク割り当てポリシー(即ち、リソースサブグラフ)の構築により、後続で、特徴抽出及び性能予測を通じて最適なタスク割り当てポリシーを選別するのに便利である。

10

【0105】

本実施例の上記のリソースサブグラフの構築プロセスは、一例に過ぎず、他のタスク割り当てポリシーの割り当て方式も本開示の実施例の保護範囲内に含まれ得る。

【0106】

いくつかの任意の実施例において、前記第1ネットワークを訓練することは、少なくとも1つのタスク割り当てポリシーの予測性能及び実際の性能に基づいて、前記第1ネットワークを訓練することを含む。

【0107】

本実施形態において、第1ネットワークを用いて各タスク割り当てポリシーの予測性能を得、予測性能と実際の性能及び誤差逆伝播方式により、前記第1ネットワークを訓練する。

20

【0108】

いくつかの任意の実施例において、少なくとも1つのタスク割り当てポリシーの予測性能を得ることは、計算グラフ及び各リソースサブグラフに基づいて、第1ネットワークを用いて各リソースサブグラフに対応する予測性能を得ることを含む。

【0109】

本実施形態において、計算グラフ及び1つのリソースサブグラフを第1ネットワークに入力することにより、前記リソースサブグラフに対応する予測性能を得る。

30

【0110】

例示的に、図5は、本開示の実施例における、予測性能の取得方式の例示的なフローチャートであり、図5に示すように、予測性能の取得方法は、以下のステップを含み得る。

【0111】

ステップ301において、前記第1ネットワークの特徴抽出モジュールを用いて前記計算グラフの特徴を抽出し、第1特徴集合を得る。

【0112】

ステップ302において、前記特徴抽出モジュールを用いて、前記少なくとも1つのリソースサブグラフの特徴をそれぞれ抽出し、少なくとも1つの第2特徴集合を得る。

【0113】

ステップ303において、前記第1特徴集合、各第2特徴集合及び前記第1ネットワークの予測モジュールに基づいて、各リソースサブグラフに対応する予測性能を得る。

40

【0114】

本実施形態において、特徴抽出モジュールを用いて計算グラフの特徴を抽出し、第1特徴集合を得る。前記第1特徴集合は、特徴セット、特徴、特徴ベクトルとも呼ばれる。グラフノード間のメッセージパッシングを通じてグラフの依存関係を捕捉するために、特徴抽出モジュールを用いて計算グラフとリソースサブグラフの特徴をそれぞれ抽出し、CPU、GPU、FPGA、DSP及びメモリなどを含み、主に、計算能力、ストレージ、通信などの次元の特徴を含む。

【0115】

50

例示的に、計算グラフの入力特徴セットを決定し、前記入力特徴セットは、入力特徴、入力特徴ベクトル、入力特徴行列とも呼ばれ、前記計算グラフ内の各ノードの入力特徴情報を含み、計算グラフの隣接行列を決定し、前記隣接行列は、前記計算グラフのトポロジ情報を示し、又は前記隣接行列は、前記計算グラフ内のノード間の関係を示し、前記入力特徴の組み合わせ、前記隣接行列及び前記特徴抽出モジュールに基づいて前記計算グラフの特徴を抽出し、第1特徴集合を得る。

【0116】

任意選択的に、前記計算グラフの入力特徴セットは、前記計算グラフ内の各ノードの特徴ベクトルを含み、前記計算グラフ内の各ノードの特徴ベクトルは、各ノードに対応する演算子を実行するために必要なリソース情報を含み、必要なリソース情報は例えば、CPU、GPU、DSP、FPGA、メモリ占有率などを含む。

10

【0117】

任意選択的に、計算グラフに対応する隣接行列内の要素は、各2つのノード間の関係の強さを示し、要素の数値は、対応する2つのノード間の伝送データに関連する。

【0118】

例示的に、図6aは、本開示の実施例における計算グラフの特徴抽出の概略図であり、図6aに示すように、計算グラフは、6個のノードを含み、それに対応して、入力特徴セットは、6組の特徴ベクトルを含み、例えば、以下の通りに表す。

【数1】

$$\begin{bmatrix} f_{11} & f_{21} & f_{31} & f_{41} & f_{51} & f_{61} \\ f_{12} & f_{22} & f_{32} & f_{42} & f_{52} & f_{61} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ f_{1M} & f_{2M} & f_{3M} & f_{4M} & f_{5M} & f_{6M} \end{bmatrix}$$

20

(1)

【0119】

ここで、各組の特徴ベクトル(上記の行列内の各列)は、計算グラフ上の1つのノード二対応し、各組の特徴ベクトルは、当該ノードに対応する演算子が実行(又は演算)されるとき各リソースの使用状況(又はリソース需要)の特徴、即ち、演算子のハードウェア実行コスト、又は演算子のハードウェア占有データを含み、例えば、CPU占有率、GPU占有率、DSP占有率、FPGA占有率、ストレージ占有率などの特徴を含んでもよい。前記占有率は、占有、占有比例、占有比率、使用、使用率、使用比例、使用比率、利用、利用率、利用比例、利用比率とも呼ばれる。計算グラフ内のノード間の接続関係により、ノード1とノード2との間に接続関係を有し、ノード2とノード3との間に接続関係を有し、ノード3とノード4との間に接続関係を有し、ノード4とノード5との間に接続関係を有し、ノード1とノード6との間に接続関係を有し、ノード6とノード4との間に接続関係を有することを決定することができ、それにより、隣接行列において対応する要素(即ち、図面に示す e_{12} 、 e_{16} 、 e_{23} 、 e_{34} 、 e_{45} 、 e_{64})は特定値を有し、他の要素の値は0であることを決定することができ、隣接行列は、以下に示すとおりである。

30

40

【数2】

$$\begin{bmatrix} 0 & e_{12} & 0 & 0 & 0 & e_{16} \\ 0 & 0 & e_{23} & 0 & 0 & 0 \\ 0 & 0 & 0 & e_{34} & 0 & 0 \\ 0 & 0 & 0 & 0 & e_{45} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & e_{64} & 0 & 0 \end{bmatrix}$$

(2)

50

【 0 1 2 0 】

ここで、上記の要素の数値は、対応する2つのノード間の伝送データに関連する。

【 0 1 2 1 】

例示的に、 $e = k d$ であり、ここで、 k は、所定係数を示し、 d は、隣接する2つのノード間の伝送データのサイズを示す。

【 0 1 2 2 】

例えば、 $e_{34} = k d_{34}$ であり、ここで、 d_{34} は、計算グラフ内のノード3とノード4との間の伝送データのサイズを示す。

【 0 1 2 3 】

すると、例えば、図に示す入力特徴セットと隣接行列を特徴抽出モジュールに入力し、それにより、第1特徴集合を得る。

10

【 0 1 2 4 】

同様に、特徴抽出モジュールを用いてリソースサブグラフの特徴を抽出して、第2特徴集合を得る。前記第2特徴集合は、特徴セット、特徴、特徴ベクトルとも呼ばれる。

【 0 1 2 5 】

例示的に、リソースサブグラフの入力特徴セットを決定し、前記入力特徴セットは、前記リソースサブグラフ内の各ノードの入力特徴情報を含み、リソースサブグラフの隣接行列を決定し、前記隣接行列は、前記リソースサブグラフのトポロジ情報を示し、又は前記隣接行列は、前記リソースサブグラフ内のノード間の関係を示し、前記入力特徴の組み合わせ、前記隣接行列及び前記特徴抽出モジュールに基づいて前記リソースサブグラフの特徴を抽出し、第2特徴集合を得る。

20

【 0 1 2 6 】

任意選択的に、前記リソースサブグラフの入力特徴セットは、前記リソースサブグラフ内の各ノードの特徴ベクトルを含み、前記リソースサブグラフ内の各ノードの特徴ベクトルは、各ノードに対応するモノのインターネット機器が有するリソース情報（能力情報とも呼ばれる）の少なくとも一部を含み、前記リソース情報の少なくとも一部は、例えば、CPU、GPU、DSP、FPGA、メモリなどの利用可能リソースを含む。

【 0 1 2 7 】

任意選択的に、リソースサブグラフに対応する隣接行列内の要素は、各2つのノード間の通信関係の強さを示し、要素の数値は、対応する2つのノード間の伝送レート及び/又は遅延等に関連する。

30

【 0 1 2 8 】

例示的に、図6bは、本開示の実施例における、リソースサブグラフの特徴抽出の概略図であり、図6bに示すように、タスク割り当てポリシー、即ち、リソースサブグラフであり、図面には、6個のノードを含み、それに対応して、入力特徴セットは、6組の特徴ベクトルを含み、例えば、以下の通りに表す。

【 数 3 】

$$\begin{bmatrix} f_{11} & f_{21} & f_{31} & f_{41} & f_{51} & f_{61} \\ f_{12} & f_{22} & f_{32} & f_{42} & f_{52} & f_{61} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ f_{1M} & f_{2M} & f_{3M} & f_{4M} & f_{5M} & f_{6M} \end{bmatrix}$$

(3)

40

【 0 1 2 9 】

ここで、各組の特徴ベクトル（上記の行列内の各列）は、リソースサブグラフ上の1つのノードに対応し、各組の特徴ベクトルは、各ノードに対応するリソースの特徴を含み、例えば、CPUリソース、GPUリソース、DSPリソース、FPGAリソース、ストレージリソースなどの特徴を含み得る。そして、計算グラフ内のノード間の接続関係により、ノードV2とノードV1との間に接続関係を有し、ノードV1とノードV3との間に接

50

続関係を有し、ノードV3とノードV4との間に接続関係を有し、ノードV3とノードV5との間に接続関係を有し、ノードV5とノードV6との間に接続関係を有することを決定することができ、それにより、隣接行列において対応する要素（即ち、図面に示す e_{13} 、 e_{21} 、 e_{34} 、 e_{35} 、 e_{56} ）は特定値を有し、他の要素の値は0であることを決定することができ、隣接行列は、以下に示すとおりである。

【数4】

$$\begin{bmatrix} 0 & 0 & e_{13} & 0 & 0 & 0 \\ e_{21} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & e_{34} & e_{35} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & e_{56} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

10

(4)

【0130】

ここで、上記の要素の数値は、対応する2つのノード間の伝送レート及び/又は遅延などに関連する。

【0131】

例示的に、 $e = k_1 t s + k_2 l$ であり、ここで、 k_1 と k_2 はそれぞれ、所定係数を示し、 $t s$ は、隣接する2つのノード間の伝送レートを示し、 l は、隣接する2つのノード間の伝送遅延を示す。

20

【0132】

例えば、 $e_{34} = k_1 t s_{34} + k_2 l_{34}$ であり、ここで、 $t s_{34}$ は、リソースサブグラフ内のノードV3とノードV4との間の伝送レートを示し、 l_{34} は、リソースサブグラフ内のノードV3とノードV4との間の遅延を示す。

【0133】

すると、例えば、図に示す入力特徴セットと隣接行列を特徴抽出モジュールに入力し、それにより、第2特徴集合を得る。

【0134】

本実施形態において、上記の計算グラフ又はリソースサブグラフに対応する入力特徴セットと隣接行列を特徴抽出モジュールの入力とし、以下の式(5)の順伝播アルゴリズムにより特徴更新を行い、特徴抽出モジュールの多層ネットワークの順伝播により、計算グラフ又はリソースサブグラフ内のすべてのノードとエッジの特徴を融合した1つの特徴集合（特徴ベクトルとも呼ばれる）を得、即ち、計算グラフに対応する第1特徴集合と、リソースサブグラフに対応する第2特徴集合を得る。

30

【数5】

$$H^{(l+1)} = \sigma(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}) \quad (5)$$

40

【0135】

ここで、

【化1】

$$\hat{A} = A + I$$

は、自己接続が追加された隣接行列を示し、 I は、単位行列を示し、

50

【化 2】

$$\hat{D}_{ii} = \sum_j \hat{A}_{ij}$$

であり、ここで、 i は、行列の行数を示し、 j は、行列の列数を示し、 $H^{(1)}$ は、特徴抽出モジュールの多層ネットワークにおける第 1 層のすべてのノード特徴を示し、 $H^{(0)}$ は、特徴抽出モジュールの入力レイヤーのすべてのノードの特徴を示し、 W は、特徴抽出モジュールの訓練可能な重み行列（即ち、ネットワークパラメータ）を示し、 $W^{(1)}$ は、特徴抽出モジュールの多層ネットワークにおける第 1 層の訓練可能な重み行列を示し、

10

【化 3】

$$\sigma(\cdot)$$

は、活性化関数を示す。

【0136】

20

いくつかの任意の実施例において、前記第 1 特徴集合、各第 2 特徴集合及び前記第 1 ネットワークの予測モジュールに基づいて、各リソースサブグラフに対応する予測性能を得ることは、前記第 1 特徴集合及び各第 2 特徴集合に基づいて、少なくとも 1 つの第 3 特徴集合を得ることであって、各第 3 特徴集合は、前記第 1 特徴集合及び各第 2 特徴集合を含む、ことと、各第 3 特徴集合及び前記予測モジュールに基づいて、各リソースサブグラフに対応する予測データを得、各リソースサブグラフに対応する予測データに基づいて、各リソースサブグラフに対応する予測性能を得ることと、を含む。

【0137】

本実施形態において、少なくとも 1 つのタスク割り当てポリシーから、最適なタスク割り当てポリシーを選択するために、本実施形態において、予測モジュールを用いて、各タスク割り当てポリシーを実行するシステム性能を予測する。異なるオペレーティングシステムでのタスクスケジューリングとリソース割り当ての本質的な統計法則を学習して、動的に変化するモノのインターネットエッジコンピューティング環境に適応するために、複雑な問題の非線形回帰予測を処理するための予測モジュールを構築し、異なるタスク割り当てポリシーとシステム性能との間の対応関係を暗黙的に含む多数の履歴サンプルからその統計的規則性を学習することができる。

30

【0138】

本実施形態において、予測モジュールの入力データは、特徴抽出モジュールを用いてそれぞれ得られた第 1 特徴集合と 1 つの第 2 特徴集合の融合特徴（即ち、第 3 特徴集合）である。例示的に、第 1 特徴集合と第 2 特徴集合を繋ぎ合わせて、第 3 特徴集合を得ることができる。更に、第 3 特徴集合を予測モジュールに入力し、予測モジュールにおける多層ネットワークの順伝播アルゴリズムの層ごとの反復によって、各リソースサブグラフに対応する予測データを得る。

40

【0139】

いくつかの任意の実施例において、前記予測データは、性能指標（鍵性能指標、システム性能指標、鍵システム性能指標とも呼ばれる）の予測値を示すことができ、前記処理対象タスクを実行する予測実行時間長と、前記処理対象タスクを実行する予測エネルギー消費量と、前記処理対象タスクを実行する予測信頼性と、のうちの少なくとも 1 つを含む。前記予測実行時間長が実行時間長の予測値であり、前記予測エネルギー消費量がエネルギー消費量の予測値であり、前記予測信頼性が信頼性の予測値である。例示的に、予測デー

50

は、3つのデータ又は3つの成分を含む1つのベクトルであってもよく、ここで、1つのデータ又は成分は、前記処理対象タスクを実行する予測実行時間長を示し、例えば、
【化4】

$$\lambda_t$$

と表記してもよく、1つのデータ又は成分は、前記処理対象タスクを実行する予測エネルギー消費量を示し、例えば、
【化5】

$$\lambda_e$$

と表記してもよく、1つのデータ又は成分は、前記処理対象タスクを実行する予測信頼性を示し、例えば、
【化6】

$$\lambda_r$$

と表記してもよく、上記の予測データに基づいて各リソースサブグラフに対応する予測性能を決定する。ここで、前記性能は、システム全体の性能とも呼ばれる。

【0140】

いくつかの任意の実施例において、前記各リソースサブグラフに対応する予測データに基づいて、各リソースサブグラフに対応する予測性能を得ることは、所定重みに従って、各リソースサブグラフに対応する予測データに対して重み付け処理を行い、各リソースサブグラフに対応する予測性能を得ることを含む。

【0141】

本実施形態において、予測データが3つの成分を含むことを例にとると、各成分に対応する所定重みに従って重み付け処理を行い、つまり、式(6)に従って対応する予測性能
【化7】

$$\eta$$

を得、ここで、
【化8】

$$Q(\cdot)$$

は関数を示し、ここで、各成分又はデータ又は(鍵)性能指標に対する重み付け情報を含む。

【数6】

$$\eta = Q(\lambda_t, \lambda_e, \lambda_r, \dots) \quad (6)$$

【0142】

式(6)の関数式の具体的な形式、即ち、所定重みの具体的な情報は、異なるシーンの遅延、エネルギー消費量、信頼性等に対する異なる要件、或いは、重要度又は注目度に依存するため、つまり、特定関数を使用して、異なる性能指標に重み付けることにより、複数の性能指標間のトレードオフを実現し、設定された式に従って各主要な性能指標の重み付けられた値を計算することで、システム全体の性能を得ることができる。つまり、式(6)によって得られた予測性能は、サービス品質(QoS: Quality of Service)に関連するシステム全体の性能を反映する。

【0143】

いくつかの任意の実施例において、前記第1ネットワークを訓練することは、各タスク割り当てポリシーの予測性能及び実際の性能に基づいて、前記特徴抽出モジュールと前記予測モジュールを訓練することを含む。

10

【0144】

本実施形態において、具体的には、各タスク割り当てポリシーの予測性能及び訓練データ集合内の実際の性能に基づいて、特徴抽出モジュールと予測モジュールのネットワークパラメータを更新し、それにより、前記特徴抽出モジュールと前記予測モジュールを訓練する。

【0145】

いくつかの任意の実施例において、前記特徴抽出モジュールと前記予測モジュールを訓練することは、各タスク割り当てポリシーの予測性能と実際の性能の間の誤差を逆伝播し、予測性能と実際の性能の間の誤差が所定条件を満たすまで、勾配降下アルゴリズムを利用して第1ネットワークの特徴抽出モジュールと予測モジュールのネットワークパラメータを更新することを含む。

20

【0146】

例示的に、前記予測性能と実際の性能との間の誤差が所定条件を満たすことは、予測性能と実際の性能との間の誤差が、所定閾値より小さいことであってもよい。

【0147】

本実施形態において、特徴抽出モジュールの多層ネットワークによって抽出された特徴に基づいて、予測モジュールを用いて、異なるタスク割り当てポリシーとシステム性能(実行時間、エネルギー消費量、信頼性)との対応関係から、複数の異機種モノのインターネット機器の異なるオペレーティングシステムでのタスクスケジューリングの本質的な統計的規則性を学習し、タスク実行前に、与えられたタスク割り当てポリシーのシステム性能の予測を実現することで、複数のリソースサブグラフに含まれる異なるタスク割り当てポリシーから、最適なシステム性能を得ることができるタスク割り当てポリシーの選択を容易にする。処理待ちの計算タスクとモノのインターネット機器の利用可能リソースとの間の最適なマッチングを実現することにより、リソースの利用を最大化し、システム全体の性能を向上させる。

30

【0148】

本開示のいくつかの任意の実施例において、前記方法は、前記訓練データ集合を更新することを更に含み、更新された前記訓練データ集合は、前記第1ネットワークを更新するために使用される。前記第1ネットワークを更新することは、前記第1ネットワークを訓練することであり、又は第1ネットワークに対して最適化を行う(即ち、第1ネットワークを最適化すること)ことである。ここで、訓練データ集合の更新は、タスク割り当てポリシー(例えば、リソースサブグラフ)の更新及び実際の性能の更新を含み得、任意選択的に、訓練データ集合は更に計算グラフを含んでもよく、すると、訓練データ集合の更新は、計算グラフの更新を更に含み得る。

40

【0149】

いくつかの任意の実施例において、前記訓練データ集合を更新することは、計算グラフ及びリソースグラフに基づいて、ヒューリスティック法、グラフ探索法、グラフ最適化方法及びサブグラフマッチング方法のうち少なくとも1つの方法を用いて、少なくとも1つのリソースサブグラフを生成し、各リソースサブグラフに対応するタスク

50

割り当てポリシーに従って実際に実行した後、各リソースサブグラフに対応する実際の性能を得、前記計算グラフ、各リソースサブグラフ及び対応する実際の性能を前記訓練データ集合に追加すること、

計算グラフ及びリソースグラフに基づいて、ヒューリスティック法、グラフ探索法、グラフ最適化方法及びサブグラフマッチング方法のうち少なくとも1つの方法を用いて、少なくとも1つのリソースサブグラフを生成し、第1ネットワークを介して各リソースサブグラフに対応する予測性能を得、前記少なくとも1つのリソースサブグラフから予測性能が最も良いリソースサブグラフを選択し、前記予測性能が最も良いリソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、実際の性能を得、前記計算グラフ、前記予測性能が最も良いリソースサブグラフ及び対応する実際の性能を前記訓練データ集合に追加すること、

10

計算グラフ及びリソースグラフに基づいて、ランダムウォーク方法を用いて、少なくとも1つのリソースサブグラフを生成し、各リソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、実際の性能を得、前記計算グラフ、少なくとも1つのリソースサブグラフ及び実際の性能を前記訓練データ集合に追加すること、のうちの少なくとも1つを含む。

【0150】

本実施形態において、持続的に更新された訓練データ集合を用いて第1ネットワークを周期的に訓練し、それにより、第1ネットワークベースのタスク割り当てシステム又はプラットフォームを、自己学習及び自己適応能力を有し、インテリジェントな自己適応を実現し、「使えば使うほど賢くなる」という効果を実現させる。

20

【0151】

ここで、訓練データ集合の更新は、少なくとも上記のいくつかの方式を含む。第1の方式において、ヒューリスティック法、グラフ探索法、グラフ最適化方法及びサブグラフマッチング方法のうち少なくとも1つの方法を採用して、少なくとも1つのタスク割り当てポリシーを得、対応する少なくとも1つのタスク割り当てポリシーに従って実際のタスク実行を実行した後、対応する実際の性能を記録し、採用された少なくとも1つのタスク割り当てポリシー及び対応する実際の性能を新しいサンプルデータとして訓練データ集合に追加する。第2の方式において、まず、ヒューリスティック法、グラフ探索法、グラフ最適化方法及びサブグラフマッチング方法のうち少なくとも1つの方法を採用して、少なくとも1つのタスク割り当てポリシーを得、第1ネットワークを用いて、そのうちの予測性能が最も良いタスク割り当てポリシーを決定し、実際に当該予測性能が最も良いタスク割り当てポリシーに従ってタスク実行を実行した後、得られた実際の性能を記録し、当該予測性能が最も良いタスク割り当てポリシー及び対応する実際の性能を新しいサンプルデータとして訓練データ集合に追加することである。第3の方式において、システムがビジー状態でないときに、リソースグラフでランダムウォークを実行して処理対象タスク演算子を割り当て、異なる割り当てポリシーを持つ複数のリソースサブグラフを生成し、各リソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後に実際の性能を得、前記タスク割り当てポリシーと実際の性能を前記訓練データ集合に追加する。それにより、貪欲なヒューリスティックに基づく単純な固定とモード制約のリソースサブグラフ構築方法が局所最適に陥りやすいという欠点を克服し、タスク割り当てポリシーの多様性を高めることにより、最適なシステム性能を得る可能性の高いタスク割り当てポリシーを得ることができる。

30

40

【0152】

任意選択的に、本開示の実施例の上記のリソースサブグラフの生成方法（即ち、前記計算グラフ内の第1ノードを決定し、前記第1ノードは、リソース需要が最も大きいノードであり、前記リソースグラフ内の少なくとも1つの第2ノードを決定し、前記少なくとも1つの第2ノードは、前記第1ノードのリソース需要を満たすノードであり、各第2ノードに基づいて1つのリソースサブグラフを決定し、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表す方法）は

50

、ヒューリスティック法であってもよく、又はグラフ探索法であってもよく、又はグラフ最適化方法であってもよく、又はサブグラフマッチング方法であってもよく、本実施例は、上記のリソースサブグラフの生成方法に限定されずタスク割り当てポリシーを生成することができ、他のヒューリスティック法、グラフ探索法、グラフ最適化方法及びサブグラフマッチング方法のうちの少なくとも1つの方法を採用してタスク割り当てポリシーを生成することもできる。

【0153】

本開示の実施例は更に、モノのインターネット機器ベースのタスク割り当て方法を提供する。図7は、本開示の実施例のモノのインターネット機器ベースのタスク割り当て方法の例示的なフローチャートであり、図7に示すように、前記方法は、以下のステップを含む。

10

【0154】

ステップ401において、処理対象タスクに対応する計算グラフ、及び1つ又は複数のモノのインターネット機器に対応するリソースグラフを決定する。

【0155】

ステップ402において、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成する。

【0156】

ステップ403において、前記少なくとも1つのタスク割り当てポリシーを第1ネットワークに入力し、各タスク割り当てポリシーに対応する予測性能を取得する。

20

【0157】

ステップ404において、予測性能が最も良いタスク割り当てポリシーを決定し、決定されたタスク割り当てポリシーに基づいてタスク割り当てを行う。

【0158】

本実施例において、第1ネットワークは、前述のネットワーク訓練方法の実施例における詳細な説明を参照して最適化され、最適化された第1ネットワークを得ることができる。

【0159】

本実施形態において、前記タスク割り当てポリシーは、処理対象タスクを少なくとも1つのモノのインターネット機器によって実行されるポリシーに割り当ててことを示し、言い換えれば、前記タスク割り当てポリシーを用いて、少なくとも1つのモノのインターネット機器を決定し、少なくとも1つのモノのインターネット機器を用いて、タスク割り当てポリシーの指示に従って処理対象タスクを実行することができる。任意選択的に、前記タスク割り当てポリシーは、タスク割り当て方法、タスク割り当て方式、タスクスケジューリングポリシー、タスクスケジューリング方法、タスクスケジューリング方式、タスク割り付けポリシー、タスク割り付け方法、タスク割り付け方式などのうちの1つとも呼ばれる。

30

【0160】

本実施例のステップ401ないしステップ404は、具体的に、前述のネットワーク訓練方法の実施例の具体的な説明を参照することができる。違いは、本実施例において、各タスク割り当てポリシーに対して、第1ネットワークを用いて対応する予測性能を得、これから、予測性能が最も良いタスク割り当てポリシーを選択してタスク割り当てを行うことである。例示的に、第1ネットワークを用いて、各タスク割り当てポリシーに対応するシステム性能の予測値を得ることができ、最大予測値に対応するタスク割り当てポリシーを選択してタスク割り当てを行う。

40

【0161】

本開示のいくつかの任意の実施例において、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成することは、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成することを含み、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表し、前記タスク割り当てポリシーは、対応するリソ

50

ースグラフの少なくとも1つのノードを前記計算グラフの各ノードに割り当てるために使用され、リソースサブグラフ内の1つのノードは、1つ又は複数のモノのインターネット機器の能力及びノ又はリソースの少なくとも一部を表し、リソースサブグラフ内の隣接する2つのノードのエッジは、1つ又は複数のモノのインターネット機器の能力及びノ又はリソースの少なくとも一部の間を表す。

【0162】

本実施例は具体的に、前述のネットワーク訓練方法の実施例における具体的な説明を参照することができ、ここでは、繰り返して説明しない。

【0163】

本開示のいくつかの任意の実施例において、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成することは、前記計算グラフ内の第1ノードを決定することであって、前記第1ノードは、リソース需要が最も大きいノードである、ことと、前記リソースグラフ内の少なくとも1つの第2ノードを決定することであって、前記少なくとも1つの第2ノードは、前記第1ノードのリソース需要を満たすノードである、ことと、各第2ノードに基づいて、1つのリソースサブグラフを決定することであって、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表す、ことと、を含む。

10

【0164】

本実施例は具体的に、前述のネットワーク訓練方法の実施例における具体的な説明（例えば、図3と図4に示す詳細な記載）を参照することができ、ここでは、繰り返して説明しない。

20

【0165】

本開示のいくつかの任意の実施例において、前記各タスク割り当てポリシーに対応する予測性能を得ることは、計算グラフ及び各リソースサブグラフに基づいて、第1ネットワークを用いて各リソースサブグラフに対応する予測性能を得ることを含む。

【0166】

本実施例は具体的に、前述のネットワーク訓練方法の実施例における具体的な説明を参照することができ、ここでは、繰り返して説明しない。

【0167】

いくつかの任意の実施例において、前記第1ネットワークを用いて各リソースサブグラフに対応する予測性能を得ることは、前記第1ネットワークの特徴抽出モジュールを用いて前記計算グラフの特徴を抽出し、第1特徴集合を得ることと、前記特徴抽出モジュールを用いて、前記少なくとも1つのリソースサブグラフの特徴をそれぞれ抽出し、少なくとも1つの第2特徴集合を得ることと、前記第1特徴集合、各第2特徴集合及び前記第1ネットワークの予測モジュールに基づいて、各リソースサブグラフに対応する予測性能を得ることと、を含む。

30

【0168】

本実施例は具体的に、前述のネットワーク訓練方法の実施例における具体的な説明（例えば、図5、図6a及び図6bに示す詳細な記載）を参照することができ、ここでは、繰り返して説明しない。

40

【0169】

いくつかの任意の実施例において、前記第1特徴集合、各第2特徴集合及び前記第1ネットワークの予測モジュールに基づいて、各リソースサブグラフに対応する予測性能を得ることは、前記第1特徴集合及び各第2特徴集合に基づいて、少なくとも1つの第3特徴集合を得ることであって、各第3特徴集合は、前記第1特徴集合及び各第2特徴集合を含む、ことと、各第3特徴集合及び前記予測モジュールに基づいて、各リソースサブグラフに対応する予測データを得、各リソースサブグラフに対応する予測データに基づいて、各リソースサブグラフに対応する予測性能を得ることと、を含む。

【0170】

本実施例は具体的に、前述のネットワーク訓練方法の実施例における具体的な説明を参

50

照することができ、ここでは、繰り返して説明しない。

【0171】

いくつかの任意の実施例において、前記予測データは、前記処理対象タスクを実行する予測実行時間長と、前記処理対象タスクを実行する予測エネルギー消費量と、前記処理対象タスクを実行する予測信頼性と、のうちの少なくとも1つを含む。

【0172】

本実施例は具体的に、前述のネットワーク訓練方法の実施例における具体的な説明を参照することができ、ここでは、繰り返して説明しない。

【0173】

いくつかの任意の実施例において、前記各リソースサブグラフに対応する予測データに基づいて、各リソースサブグラフに対応する予測性能を得ることは、所定重みに従って、各リソースサブグラフに対応する予測データに対して重み付け処理を行い、各リソースサブグラフに対応する予測性能を得ることを含む。

10

【0174】

本実施例は具体的に、前述のネットワーク訓練方法の実施例における具体的な説明を参照することができ、ここでは、繰り返して説明しない。

【0175】

いくつかの任意の実施例において、前記予測性能が最も良いタスク割り当てポリシーを決定し、決定されたタスク割り当てポリシーに基づいてタスク割り当てを行うことは、各リソースサブグラフに対応する予測性能に基づいて、予測性能（即ち、システム全体の性能指標の予測値）の数値が最も大きい1つの対応するタスク割り当てポリシーを選択し、実際に当該ポリシーに従ってタスク割り当てを行うことを含む。

20

【0176】

本開示のいくつかの任意の実施例において、前記方法は、タスク割り当てを行った後、前記処理対象タスクが対応するタスク割り当てポリシーに従って実行されるときの実際の性能を取得し、対応するタスク割り当てポリシー及び取得された実際の性能を訓練データ集合に記憶することを更に含み、前記訓練データ集合は、前記第1ネットワークを更新するために使用される。

【0177】

本実施形態において、上記の決定された予測性能が最も良いタスク割り当てポリシーに従ってタスク割り当てを行った後、当該タスク割り当てポリシーに従って処理対象タスクを実際に行って、実際の性能（又は実際のシステム全体の性能）を得、対応するタスク割り当てポリシー及び取得された実際の性能を、第1ネットワークを更新するための訓練データ集合に記憶して、訓練データ集合を構成するか、又は訓練データ集合を更新する。

30

【0178】

以下では、1つの特定例を参照して、本開示の実施例のタスク割り当てについて説明する。

【0179】

図8は、本開示の実施例のモノのインターネット機器ベースのタスク割り当て方法の概略図であり、図8に示すように、ステップ1において、1つの処理対象タスクに対して、処理対象タスクの計算グラフを決定し、任意選択的に、計算グラフを最適化することができ、例えば、いくつかのノードをマージして、最適化された計算グラフを得る。なお、本開示の上記の実施例における計算グラフは、最適化された計算グラフを指すことができる。更に、特定のルールに従って、計算グラフ内の各ノードに番号を付ける。

40

【0180】

ステップ2において、計算グラフと、システムにおける各モノのインターネット機器のリソースと能力状況に基づいて構築されたリソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成する。ここで、リソースサブグラフの生成方式は、前述の実施例の記載を参照することができ、ここでは、繰り返して説明しない。そして、図8に示すリソースサブグラフは、具体的に、図4のリソースサブグラフを参照することができる。

50

【 0 1 8 1 】

ステップ 3 において、計算グラフとリソースサブグラフに対してそれぞれ前処理を行い、具体的には、計算グラフに対応する入力特徴セット（入力特徴、又は入力特徴行列とも呼ばれる）及び隣接行列を決定し、及び各リソースサブグラフに対応する入力特徴セット（入力特徴、又は入力特徴行列とも呼ばれる）及び隣接行列を決定することであり、更に、計算グラフに対応する入力特徴セット（入力特徴、又は入力特徴行列とも呼ばれる）及び隣接行列を特徴抽出モジュールにそれぞれ入力して特徴抽出を行い、第 1 特徴集合を得る。各リソースサブグラフに対応する入力特徴セット（入力特徴、又は入力特徴行列とも呼ばれる）及び隣接行列を特徴抽出モジュールに入力して特徴抽出を行い、第 2 特徴集合を得る。ここで、例示的に、特徴抽出モジュールは、グラフ畳み込みニューラルネットワーク（GCN）によって実現されることができる。

10

【 0 1 8 2 】

ステップ 4 において、第 1 特徴集合と第 2 特徴集合は、第 3 特徴集合に融合され、第 3 特徴集合を予測モジュールに入力してシステム性能の予測を行い、各リソースサブグラフに対応する予測性能を得る。ここで、例示的に、前記予測モジュールは、深層ニューラルネットワーク（DNN）によって実現されることができる。

【 0 1 8 3 】

ここで、予測モジュールを用いて、例えば、少なくとも前記処理対象タスクを実行する予測実行時間長、前記処理対象タスクを実行する予測エネルギー消費量、前記処理対象タスクを実行する予測信頼性

20

【 化 9 】

$$\lambda_r$$

などの性能指標を含む予測データを得、所定重みを用いて、これらの予測データに対して重み付け処理を行い、各リソースサブグラフに対応する予測性能

【 化 1 0 】

$$\eta$$

30

を得ることができる。更に、得られた各リソースサブグラフに対応する予測性能から

【 化 1 1 】

$$\eta$$

の値が最も大きいリソースサブグラフを選択し、それが示すタスク割り当てポリシーに基づいてタスク割り当てを行う。

【 0 1 8 4 】

任意選択的に、タスクを実際に行うときに得られた実際の性能を記録し、予測性能を実際の性能と比較して、予測性能と実際の性能との間の誤差を決定し、誤差逆伝播と勾配降下アルゴリズムを用いて、第 1 ネットワークに含まれる特徴抽出モジュールと予測モジュールのネットワークパラメータを更新し、それにより、特徴抽出モジュールと予測モジュールの訓練を実現する。

40

【 0 1 8 5 】

図 9 は、本開示の実施例のタスク割り当てシステム又はプラットフォームの構成の概略図であり、図 9 に示すように、本実施例のタスク割り当てシステムは、訓練データ集合の構築、訓練段階、推論段階及び持続学習のいくつかの部分を含む。

【 0 1 8 6 】

訓練データ集合の構築：異なる処理対象タスクによって構築された計算グラフと、複数

50

のモノのインターネット機器のリソースと能力状況によって構築されたリソースグラフを入力とし、リソースサブグラフ構築モジュールを用いて、異なるタスク割り当てポリシーを含む複数のリソースサブグラフを構築し、それをモノのインターネット機器で実際に配置し、タスクが実行された後、対応する実際の性能を記録し、各組のタスク割り当てポリシー（即ち、リソースサブグラフ）及び得られた実際の性能を訓練データとして使用し、それにより、初期訓練データ集合の構築を完了する。実際の場合、訓練データ集合は更に、各処理対象タスクに対応する計算グラフを含む。

【0187】

訓練段階：入力は、訓練データ集合内のすべての訓練データ（訓練サンプルとも呼ばれる）であり、ここで、各訓練データは、計算グラフ、リソースサブグラフ、対応する実際のシステム性能を含む。訓練データを本開示の実施例のネットワークモデルに入力し、勾配降下アルゴリズムを用いて、得られたシステム性能指標予測値 p と、システム性能の実際の値 t との間の誤差を逆伝播して、収束するまで、ネットワークモデルにおける特徴抽出モジュールと予測モジュールのネットワークパラメータ（例えば、重み）を更新し、最終的に得られたネットワークパラメータ（モデルパラメータとも呼ばれる）は、許容誤差範囲内になり（アルゴリズムで手動で設定できる）、訓練サンプルのシステム性能指標予測値を実際の値に最も近づけることができる。

10

【0188】

推論段階：訓練済みのネットワークモデルにしたって、処理対象タスクのリソースに対する需要とモノのインターネット機器によって提供できる能力又はリソース状況に基づいて、予測性能が最も良いタスク割り当てポリシーを得、当該ポリシーに従って、モノのインターネット機器で処理対象タスクに対して実際のタスク割り当てを行う。入力データは、実行対象となるタスクによって構築された計算グラフ、及びフリーリソースを有するモノのインターネット機器によって構築されたリソースグラフである。計算能力、記憶と通信の暗黙情報を大量に含む2つのグラフ構造（即ち、計算グラフとリソースグラフ）をネットワークモデルに入力し、そのうちのリソースサブグラフ構築、特徴抽出及び性能予測の3つのモジュールを用いて、各リソースサブグラフ（即ち、異なるタスク割り当てポリシー）に対応するシステム性能予測値を得る。システム性能予測値が最も大きい（即ち、 p が最も大きい）タスク割り当てポリシーを選択して最適なタスク割り当てポリシーとして使用し、このタスク割り当てポリシーに従って、実際の実行プロセスで、処理対象タスクの演算子に対応するモノのインターネット機器に配置して、タスクの配置と実行を完了する。

20

30

【0189】

持続学習段階：持続学習メカニズムの実現は、持続的に更新された訓練データ集合を用いて、ネットワークモデルにおける特徴抽出モジュールと予測モジュールのネットワークパラメータを周期的に訓練し、それにより、タスク割り当てシステム、プラットフォームを、自己学習と自己適応能力を有し、インテリジェントな自己適応を実現し、「使えば使うほど賢くなる」という効果を実現させる。具体的には、履歴サンプルの蓄積とランダムウォークの方式によって実現される。履歴サンプルの蓄積は、計算タスクを実際に行うたびに、それに採用されるタスク割り当てポリシー及び得られた実際のシステム性能を記録し、新しい訓練サンプルとして訓練データ集合に記憶する。ランダムウォークの具体的な実施方法は、上記の記載の通りである。

40

【0190】

本開示の以上の各実施例は、モノのインターネットの分散エッジコンピューティングシステムにおいて効率的な深層学習を可能にするインテリジェントコンピューティングタスク割り当て方法（ICTA：Intelligent Computing Task Allocation）を提案し、ICTAに基づいて、異機種モノのインターネット機器間の深層学習タスクのインテリジェントな割り当てシステム、プラットフォームを構築する。ICTAは主に、リソースサブグラフ構築、特徴抽出、性能予測を含む。入力は、現在深層学習タスクによって構築された計算グラフと、フリーリソースを有するモノのイン

50

ターネットエッジ機器によって構築されたリソースグラフであり、両方に基づいて、グラフ検索やサブグラフマッチングなどの方法を用いてリソースサブグラフを構築して、異なるタスク割り当てポリシーを含む複数のリソースサブグラフを生成し、モノのインターネット機器上の利用可能リソースの最大限の活用及び処理対象タスクの演算子レベルの割り当てと最適化を実現する。多層ニューラルネットワークを用いて、特徴抽出と性能予測をそれぞれ行い、リソースサブグラフと計算グラフを特徴抽出モジュールにそれぞれ入力して特徴抽出と融合を行い、2つのタイプのグラフにおいてノード及びグラフトポロジに隠された、計算能力、ストレージ及び通信などの次元に関する特徴を最大限に探索する。そして、融合された特徴を性能予測モジュールに入力してシステム性能の予測を行い、特徴抽出と性能予測モジュールのエンドツーエンド (end-to-end) の訓練により、異なるタスク割り当てポリシーとシステム性能との対応関係、及び異なるオペレーティングシステム上のタスクスケジューリングの本質的な統計的規則性を学習し、タスクを実際に実行する前に、与えられたタスク割り当てポリシーの正確なシステム性能予測を実現して、代替案から、最適な割り当てポリシーを選択し、計算タスクと利用可能リソースとの間の最適なマッチングを実現し、それにより、リソース利用率を最大化し、システム全体の性能を向上させる。持続学習メカニズムを導入し、継続的に更新される訓練データ集合を用いてICTAに対して周期的な訓練を行い、システム性能オッ呼び動的に変化される環境への適応性を更に向上させ、自己適応と自己学習の能力を有するようにし、インテリジェントな自己適応を実現し、タスク割り当てシステムに「使えば使うほど賢くなる」という効果を実現させる。

10

20

【0191】

具体的に、本開示及び以上の各実施例は、以下の技術的要点及び利点を有する。

【0192】

1. インテリジェントなタスク割り当て方法に基づく異機種モノのインターネット機器間の深層学習タスクのインテリジェントな割り当てシステム、プラットフォームを構築し、訓練データ集合の構築段階、訓練段階、推論段階、持続学習段階を含む。異機種モノのインターネット機器間の深層学習モデルの分散訓練と推論のための構築アイデアを提供し、エンドツーエンドの自動的に最適化されたモノのインターネット間の異機種機器の分散エッジコンピューティング生態モードの生成を容易にする。

【0193】

2. インテリジェントな計算タスク割り当て方法 (ICTA) を提案し、モノのインターネットの分散エッジコンピューティングシステムにおいて効率的な深層学習を可能にし、リソースサブグラフの構築、特徴抽出、性能予測などを含み、異機種モノのインターネット機器間の高性能かつインテリジェントに自己適応する深層学習計算タスクの最適な割り当てを実現する。

30

【0194】

3. 特徴抽出モジュールは、リソースサブグラフと計算グラフのノードとトポロジ特徴をそれぞれ抽出し、特徴融合を行う。深層学習計算タスクの性能に重要な役割を果たす、計算能力、記憶、通信などの次元における特徴の深層感知、特徴抽出及び特徴マッチングを実現する。

40

【0195】

4. 性能予測は、融合特徴に基づいて、多層ニューラルネットワークを用いて、異なるオペレーティングシステム上のタスクスケジューリングの本質的な統計的規則性を学習し、特徴抽出と性能予測モジュールのエンドツーエンドな訓練により、異なるタスク割り当てポリシーとシステム性能との間の対応関係を探索し、システム性能を予測して、代替案から、最適なシステム性能 (予測) に対応するタスク割り当てポリシーを選択して、深層学習タスクの実際の実行を容易にする。深層学習計算タスクとモノのインターネット機器上の利用可能リソースとの間の最適なマッチングを実現し、それにより、リソースの活用を最大化し、システム性能を向上させる。

【0196】

50

5. 持続学習メカニズムの実現は主に、履歴サンプルの蓄積、ランダムウォークの2つの方式があり、この2つの方式に基づいて、訓練集合を継続的に更新し、特徴抽出モジュールと予測モジュールを周期的に訓練して、システム性能を向上させ、環境の動的変化に適応し、自己適応と自己学習の能力を有するようにし、インテリジェントな自己適応を実現し、「使えば使うほど賢くなる」という効果を実現させる。

【0197】

本開示の実施例は、1つ又は複数のモノのインターネット機器ベースのネットワーク訓練装置を提供する。図10は、本開示の実施例のモノのインターネット機器ベースのネットワーク訓練装置の構成の例示的な構造図1であり、図10に示すように、前記装置は、第1決定ユニット11と、訓練ユニット12とを備え、ここで、

10

前記第1決定ユニット11は、訓練データ集合を決定するように構成され、前記訓練データ集合は、少なくとも1つのタスク割り当てポリシー及び対応する実際の性能を含み、1つの実際の性能は、対応するタスク割り当てポリシーに基づいて実際に実行することによって得られ、

前記訓練ユニット12は、前記訓練データ集合に基づいて第1ネットワークを訓練するように構成され、前記第1ネットワークは、タスク割り当てポリシーの性能を予測するために使用される。

【0198】

本開示のいくつかの任意の実施例において、図11に示すように、前記装置は更に、処理対象タスクに対応する計算グラフ、及び1つ又は複数のモノのインターネット機器に対応するリソースグラフを決定し、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成するように構成される、第1生成ユニット13を備える。

20

【0199】

本開示のいくつかの任意の実施例において、前記第1生成ユニット13は、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成するように構成され、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表し、前記タスク割り当てポリシーは、対応するリソースグラフの少なくとも1つのノードを前記計算グラフの各ノードに割り当てるために使用され、リソースサブグラフ内の1つのノードは、1つ又は複数のモノのインターネット機器の能力及び/又はリソースの少なくとも一部を表し、リソースサブグラフ内の隣接する2つのノードのエッジは、1つ又は複数のモノのインターネット機器の能力及び/又はリソースの少なくとも一部の間の関係を表す。

30

【0200】

本開示のいくつかの任意の実施例において、前記第1生成ユニット13は、前記計算グラフ内の第1ノードを決定し、前記第1ノードは、リソース需要が最も大きいノードであり、前記リソースグラフ内の少なくとも1つの第2ノードを決定し、前記少なくとも1つの第2ノードは、前記第1ノードのリソース需要を満たすノードであり、各第2ノードに基づいて、1つのリソースサブグラフを決定し、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表すように構成される。

40

【0201】

本開示のいくつかの任意の実施例において、前記訓練ユニット12は、少なくとも1つのタスク割り当てポリシーの予測性能及び実際の性能に基づいて、前記第1ネットワークを訓練するように構成される。

【0202】

本開示のいくつかの任意の実施例において、前記訓練ユニット12は更に、計算グラフ及び各リソースサブグラフに基づいて、第1ネットワークを用いて各リソースサブグラフに対応する予測性能を得るように構成される。

【0203】

50

本開示のいくつかの任意の実施例において、前記訓練ユニット12は、前記第1ネットワークの特徴抽出モジュールを用いて前記計算グラフの特徴を抽出し、第1特徴集合を得、前記特徴抽出モジュールを用いて、前記少なくとも1つのリソースサブグラフの特徴をそれぞれ抽出し、少なくとも1つの第2特徴集合を得、前記第1特徴集合、各第2特徴集合及び前記第1ネットワークの予測モジュールに基づいて、各リソースサブグラフに対応する予測性能を得るように構成される。

【0204】

本開示のいくつかの任意の実施例において、前記訓練ユニット12は、前記第1特徴集合及び各第2特徴集合に基づいて、少なくとも1つの第3特徴集合を得、各第3特徴集合は、前記第1特徴集合及び各第2特徴集合を含み、各第3特徴集合及び前記予測モジュールに基づいて、各リソースサブグラフに対応する予測データを得、各リソースサブグラフに対応する予測データに基づいて、各リソースサブグラフに対応する予測性能を得るように構成される。

10

【0205】

本開示のいくつかの任意の実施例において、前記予測データは、
前記処理対象タスクを実行するための予測実行時間長と、
前記処理対象タスクを実行するための予測エネルギー消費量と、
前記処理対象タスクを実行するための予測信頼性と、のうちの少なくとも1つを含む。

【0206】

本開示のいくつかの任意の実施例において、前記訓練ユニット12は、所定重みに従って、各リソースサブグラフに対応する予測データに対して重み付け処理を行い、各リソースサブグラフに対応する予測性能を得るように構成される。

20

【0207】

本開示のいくつかの任意の実施例において、前記訓練ユニット12は、各タスク割り当てポリシーの予測性能及び実際の性能に基づいて、前記特徴抽出モジュールと前記予測モジュールを訓練するように構成される。

【0208】

本開示のいくつかの任意の実施例において、前記訓練ユニット12は、各タスク割り当てポリシーの予測性能と実際の性能の間の誤差を逆伝播し、予測性能と実際の性能の間の誤差が所定条件を満たすまで、勾配降下アルゴリズムを利用して第1ネットワークの特徴抽出モジュールと予測モジュールのネットワークパラメータを更新するように構成される。

30

【0209】

本開示のいくつかの任意の実施例において、図12に示すように、前記装置は更に、前記訓練データ集合を更新するように構成される更新ユニット14を備え、更新された前記訓練データ集合は、前記第1ネットワークを更新(訓練とも呼ばれる)するために使用される。

【0210】

本開示のいくつかの任意の実施例において、前記更新ユニット14は、
計算グラフ及びリソースグラフに基づいて、ヒューリスティック法、グラフ探索法、グラフ最適化方法及びサブグラフマッチング方法のうちの少なくとも1つの方法を用いて、
少なくとも1つのリソースサブグラフを生成し、各リソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、各リソースサブグラフに対応する実際の性能を得、前記計算グラフ、各リソースサブグラフ及び対応する実際の性能を前記訓練データ集合に追加する方式、

40

計算グラフ及びリソースグラフに基づいて、ヒューリスティック法、グラフ探索法、グラフ最適化方法及びサブグラフマッチング方法のうちの少なくとも1つの方法を用いて、
少なくとも1つのリソースサブグラフを生成し、第1ネットワークを介して各リソースサブグラフに対応する予測性能を得、前記少なくとも1つのリソースサブグラフから予測性能が最も良いリソースサブグラフを選択し、前記予測性能が最も良いリソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、実際の性能を得、前記計

50

算グラフ、前記予測性能が最も良いリソースサブグラフ及び対応する実際の性能を前記訓練データ集合に追加する方式、

計算グラフ及びリソースグラフに基づいて、ランダムウォーク方法を用いて、少なくとも1つのリソースサブグラフを生成し、各リソースサブグラフに対応するタスク割り当てポリシーに従って実際に実行した後、実際の性能を得、前記計算グラフ、少なくとも1つのリソースサブグラフ及び実際の性能を前記訓練データ集合に追加する方式、の少なくとも1つの方式を用いて、前記訓練データ集合を更新するように構成される。

【0211】

ここで、本開示の実施例において前記リソースサブグラフの生成方法は、ヒューリスティック法であってもよく、又はグラフ探索法であってもよく、又はサブグラフマッチング方法であってもよく、本実施例は、上記のリソースサブグラフの生成方法に限定されずタスク割り当てポリシーを生成することができ、他のヒューリスティック法、グラフ探索法及びサブグラフマッチング方法のうちの少なくとも1つの方法を採用してタスク割り当てポリシーを生成することもできる。

【0212】

本開示の実施例において、前記装置における第1決定ユニット11、訓練ユニット12、第1生成ユニット13及び更新ユニット14は、実際の適用において、CPU、GPU、DSP、マイクロコントローラユニット(MCU: Microcontroller Unit)又はFPGA、TPU、ASIC、又はAIチップなどによって実現されることができ、

【0213】

なお、上記の実施例によるモノのインターネット機器ベースのネットワーク訓練装置がネットワーク訓練を行うとき、上記の各プログラムモジュールの分割のみを例に挙げて説明したが、実際の応用では、必要に応じて上記の処理を異なるプログラムモジュールにより割り当てられて完了してもよく、即ち、装置の内部構造を異なるプログラムモジュールに分割して、上述の処理のすべて又は一部を完了する。更に、上記の実施例によるモノのインターネット機器ベースのネットワーク訓練装置は、1つ又は複数のモノのインターネット機器ベースのネットワーク訓練方法の実施例と同じ構想に属し、その具体的な実現プロセスについては方法の実施例を参照し、ここでは繰り返して説明しない。

【0214】

本開示の実施例は更に、モノのインターネット機器ベースのタスク割り当て装置を提供する。図13は、本開示の実施例のモノのインターネット機器ベースのタスク割り当て装置の構成の例示的な構造図であり、図13に示すように、前記装置は、第2決定ユニット21、第2生成ユニット22、予測ユニット23及びタスク割り当てユニット24を備え、ここで、

前記第2決定ユニット21は、処理対象タスクに対応する計算グラフ、及び1つ又は複数のモノのインターネット機器に対応するリソースグラフを決定するように構成され、

前記第2生成ユニット22は、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのタスク割り当てポリシーを生成するように構成され、

前記予測ユニット23は、前記少なくとも1つのタスク割り当てポリシーを第1ネットワークに入力し、各タスク割り当てポリシーに対応する予測性能を取得するように構成され、

前記タスク割り当てユニット24は、予測性能が最も良いタスク割り当てポリシーを決定し、決定されたタスク割り当てポリシーに基づいてタスク割り当てを行うように構成される。

【0215】

本開示のいくつかの任意の実施例において、前記第2生成ユニット22は、前記計算グラフ及び前記リソースグラフに基づいて、少なくとも1つのリソースサブグラフを生成するように構成され、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表し、前記タスク割り当てポリシーは、対応

10

20

30

40

50

するリソースグラフの少なくとも1つのノードを前記計算グラフの各ノードに割り当てるために使用され、リソースサブグラフ内の1つのノードは、1つ又は複数のモノのインターネット機器の能力及びノ又はリソースの少なくとも一部を表し、リソースサブグラフ内の隣接する2つのノードのエッジは、1つ又は複数のモノのインターネット機器の能力及びノ又はリソースの少なくとも一部の間を関係を表す。

【0216】

本開示のいくつかの任意の実施例において、前記第2生成ユニット22は、前記計算グラフ内の第1ノードを決定し、前記第1ノードは、リソース需要が最も大きいノードであり、前記リソースグラフ内の少なくとも1つの第2ノードを決定し、前記少なくとも1つの第2ノードは、前記第1ノードのリソース需要を満たすノードであり、各第2ノードに基づいて、1つのリソースサブグラフを決定し、1つのリソースサブグラフは、1つのタスク割り当てポリシーを含むか、または1つのタスク割り当てポリシーを表すように構成される。

10

【0217】

本開示のいくつかの任意の実施例において、前記第1ネットワークは、本開示の上記の実施例に記載のネットワーク訓練装置によって最適化される。

【0218】

本開示のいくつかの任意の実施例において、前記予測ユニット23は、計算グラフ及び各リソースサブグラフに基づいて、第1ネットワークを用いて各リソースサブグラフに対応する予測性能を得るように構成される。

20

【0219】

本開示のいくつかの任意の実施例において、前記予測ユニット23は、前記第1ネットワークの特徴抽出モジュールを用いて前記計算グラフの特徴を抽出し、第1特徴集合を得、前記特徴抽出モジュールを用いて、前記少なくとも1つのリソースサブグラフの特徴をそれぞれ抽出し、少なくとも1つの第2特徴集合を得、前記第1特徴集合、各第2特徴集合及び前記第1ネットワークの予測モジュールに基づいて、各リソースサブグラフに対応する予測性能を得るように構成される。

【0220】

本開示のいくつかの任意の実施例において、前記予測ユニット23は、前記第1特徴集合及び各第2特徴集合に基づいて、少なくとも1つの第3特徴集合を得、各第3特徴集合は、前記第1特徴集合及び各第2特徴集合を含み、各第3特徴集合及び前記予測モジュールに基づいて、各リソースサブグラフに対応する予測データを得、各リソースサブグラフに対応する予測データに基づいて、各リソースサブグラフに対応する予測性能を得るように構成される。

30

【0221】

本開示のいくつかの任意の実施例において、前記予測データは、
前記処理対象タスクを実行するための予測実行時間長と、
前記処理対象タスクを実行するための予測エネルギー消費量と、
前記処理対象タスクを実行するための予測信頼性と、
のうちの少なくとも1つを含む。

【0222】

本開示のいくつかの任意の実施例において、前記予測ユニット23は、所定重みに従って、各リソースサブグラフに対応する予測データに対して重み付け処理を行い、各リソースサブグラフに対応する予測性能を得るように構成される。

40

【0223】

本開示のいくつかの任意の実施例において、前記タスク割り当てユニット24は、予測性能が最も良いタスク割り当てポリシーを決定し、当該ポリシーに基づいて、実際にタスク割り当てを行い実行するように構成される。

【0224】

本開示のいくつかの任意の実施例において、前記装置は更に、タスク割り当てを行った後、前記処理対象タスクが対応するタスク割り当てポリシーに従って実行されるときの実

50

際の性能を取得し、対応するタスク割り当てポリシー及び取得された実際の性能を訓練データ集合に記憶するように構成される取得ユニットを備え、前記訓練データ集合は、前記第1ネットワークを更新するために使用される。

【0225】

本開示の実施例において、前記装置における第2決定ユニット21、第2生成ユニット22、予測ユニット23、タスク割り当てユニット24及び取得ユニットは、実際の適用において、CPU、GPU、DSP、ASIC、AIチップ、MCU又はFPGAなどによって実現されることができる。

【0226】

なお、上記の実施例によるモノのインターネット機器ベースのタスク割り当て装置がタスク割り当てを行うとき、上記の各プログラムモジュールの分割のみを例に挙げて説明したが、実際の応用では、必要に応じて上記の処理を異なるプログラムモジュールにより割り当てられて完了してもよく、即ち、装置の内部構造を異なるプログラムモジュールに分割して、上述の処理のすべて又は一部を完了する。更に、上記の実施例によるモノのインターネット機器ベースのタスク割り当て装置は、1つ又は複数のモノのインターネット機器ベースのタスク割り当て方法の実施例と同じ構想に属し、その具体的な実現プロセスについては方法の実施例を参照し、ここでは繰り返して説明しない。

【0227】

本開示の実施例は更に、電子機器を提供する。図14は、本開示の実施例の電子機器のハードウェアの構成の例示的な構造図であり、図14に示すように、前記電子機器は、メモリ32と、プロセッサ31と、メモリ32に記憶されプロセッサ31で実行可能なコンピュータプログラムと、を備え、前記プロセッサ31は、前記プログラムを実行するとき、本開示の上記の実施例に記載のネットワーク訓練方法を実現し、又は、前記プロセッサは、前記プログラムを実行するとき、本開示の上記の実施例に記載のタスク割り当て方法を実現する。

【0228】

電子機器内の各コンポーネントは、バスシステム33を介して結合されることを理解することができる。理解できることとして、バスシステム33は、これらのコンポーネント間の接続通信を実現するために使用される。データバスに加えて、バスシステム33は、電力バス、制御バス及びステータス信号バスを含む。しかしながら、説明を明確にするために、図14には様々なバスをバスシステム33として表記されている。

【0229】

メモリ32は、揮発性メモリ又は不揮発性メモリであってもよく、揮発性及び不揮発性メモリの両方を含んでもよいことを理解することができる。ここで、不揮発性メモリは、読み取り専用メモリ(ROM: Read Only Memory)、プログラマブル読み取り専用メモリ(PROM: Programmable Read-Only Memory)、消去可能プログラマブル読み取り専用メモリ(EPROM: Erasable Programmable Read-Only Memory)、電氣的消去可能プログラマブル読み取り専用メモリ(EEPROM: Electrically Erasable Programmable Read-Only Memory)、磁気ランダムアクセスメモリ(FRAM(登録商標): ferromagnetic random access memory)、フラッシュメモリ(Flash Memory)、磁気表面メモリ、光ディスク、又はリードオンリーメモリ(CD-ROM: Compact Disc Read-Only Memory)であり得、磁気表面メモリは、磁気ディスクメモリ又は磁気テープメモリであり得る。揮発性メモリは、外部キャッシュとして使用される、ランダムアクセスメモリ(RAM: Random Access Memory)であり得る。例示的であるが限定的ではない例示によれば、多くの形のRAM、例えば、スタティックランダムアクセスメモリ(SRAM: Static Random Access Memory)、同期スタティックランダムアクセスメモリ(SSRAM: Synchronous Static Random Access Memory)、ダイナミックランダ

10

20

30

40

50

ムアクセスメモリ (DRAM: Dynamic Random Access Memory)、同期ダイナミックランダムアクセスメモリ (SDRAM: Synchronous Dynamic Random Access Memory)、ダブルデータレートの同期ダイナミックランダムアクセスメモリ (DDRSDRAM: Double Data Rate Synchronous Dynamic Random Access Memory)、拡張型同期ダイナミックランダムアクセスメモリ (ESDRAM: Enhanced Synchronous Dynamic Random Access Memory)、同期接続ダイナミックランダムアクセスメモリ (SLDRAM: SyncLink Dynamic Random Access Memory)、ダイレクトメモリバスランダムアクセスメモリ (DRRAM: Direct Rambus Random Access Memory) などが利用可能である。本開示の実施例で説明されるメモリ32は、これら及び任意の他の適切なタイプのメモリを含むが、これらに限定されないことを意図する。

10

【0230】

上記の本開示の実施例で開示された方法は、プロセッサ31に適用されるか、又はプロセッサ31によって実現されることができる。プロセッサ31は、信号の処理能力を有する集積回路チップであり得る。実現プロセスにおいて、上記の方法の各ステップは、プロセッサ31内のハードウェアの集積論理回路又はソフトウェアの形の命令によって完了できる。前記プロセッサ31は、汎用プロセッサ、DSP、又は他のプログラマブルロジックデバイス、ディスクリートゲート又はトランジスタロジックデバイス、ディスクリートハードウェアコンポーネントなどであり得る。プロセッサ31は、本開示の実施例における開示された各方法、ステップ及び論理ブロック図を実現又は実行できる。汎用プロセッサは、マイクロプロセッサ、又は任意の従来プロセッサなどであり得る。本開示の実施例で開示される方法のステップは、ハードウェア復号化プロセッサによって直接実行されてもよく、復号化プロセッサ内のハードウェアとソフトウェアモジュールの組み合わせによって実行されてもよい。ソフトウェアモジュールは、記憶媒体に配置でき、当該記憶媒体は、メモリ32に配置され、プロセッサ31は、メモリ32内の情報を読み取り、そのハードウェアと組み合わせて上記の方法のステップを完了する。

20

【0231】

例示的な実施例において、電子機器は、1つ又は複数の特定用途向け集積回路 (ASIC: Application Specific Integrated Circuit)、DSP、プログラマブルロジックデバイス (PLD: Programmable Logic Device)、複合プログラマブルロジックデバイス (CPLD: Complex Programmable Logic Device)、FPGA、汎用プロセッサ、コントローラ、MCU、マイクロプロセッサ (Microprocessor)、又は他の電子素子で実現でき、上記の方法を実行するために使用されることができる。

30

【0232】

例示的な実施例において、本開示の実施例は更に、コンピュータプログラムを含むメモリ32などの、コンピュータ可読記憶媒体を提供し、上記のコンピュータプログラムは、電子機器のプロセッサ31によって実行されて、上記の方法を完成することができる。コンピュータ可読記憶媒体は、FRAM (登録商標)、ROM、PROM、EPROM、EEPROM、Flash Memory、磁気表面メモリ、光ディスク、又はCD-ROMなどのメモリであってもよく、上記のメモリのうちの1つを含むか任意に組み合わせられた様々な機器であってもよい。

40

【0233】

本開示の実施例は更に、プロセッサによって実行されるとき、前記プロセッサに本開示の上記の実施例に記載のネットワーク訓練方法を実現させ、又は、プロセッサによって実行されるとき、前記プロセッサに本開示の上記の実施例に記載のタスク割り当て方法を実現させるためのコンピュータプログラムが記憶された、コンピュータ可読記憶媒体を提供する。

50

【0234】

本発明によるいくつかの方法の実施例に開示される方法は、競合することなく任意に組み合わせ、新たな方法の実施例を取得することができる。

【0235】

本発明で提供されるいくつかの製品の実施例で開示された特徴は、競合することなく任意に組み合わせ、新たな製品の実施例を取得することができる。

【0236】

本発明によるいくつかの方法、又は機器の実施例に開示される特徴は、競合することなく任意に組み合わせ、新たな方法の実施例又は機器の実施例を取得することができる。

【0237】

本開示の明細書の実施例、特許請求の範囲及び上記の添付図面における「第1」、「第2」及び「第3」等の用語は、特定の順番又は先後順序を限定するためのものではなく、類似する対象を区別するためのものであることに留意されたい。また、「含む」、「有する」という用語、及びこれらの任意の変形は、非排他的包含を網羅することが意図され、例えば、一連のステップ又はユニットを含む。方法、システム、製品又はデバイスは、明確に列挙されたこれらのステップ又はユニットに必ずしも限定されず、明確に列挙されない他のステップ又はユニット、又はこれらのプロセス、方法、製品又はデバイスに固有の他のステップ又はユニットを含み得る。

10

【0238】

本発明で提供するいくつかの実施例において、開示された機器及び方法は、他の方式で実現されてもよいことを理解されたい。上記で説明された装置の実施例は、例示的なものに過ぎず、例えば、前記ユニットの分割は、論理機能の分割に過ぎず、実際の実現では、他の分割方法を採用することができ、例えば、複数のユニット又はコンポーネントを組み合わせるか又は別のシステムに統合してもよく、その一部の特徴を無視するか実行しなくてもよい。更に、表示又は議論された各構成要素の相互結合又は直接結合又は通信接続は、いくつかのインターフェースを使用して実現することができ、機器又はユニット間の間接結合又は通信接続は、電氣的又は機械的な形であってもよく、他の形であってもよい。

20

【0239】

上記の別個の部品として説明されたユニットは、物理的に分離されていてもいなくてもよく、ユニットとして表示された部品は、物理的なユニットであってもなくてもよく、すなわち、1箇所に配置されてもよく、複数のネットワークユニットに分散されてもよく、実際の必要に応じてそのうちの一部又はすべてのユニットを選択して本実施例の技術案の目的を実現することができる。

30

【0240】

更に、本発明の各実施例における各機能ユニットは、すべて1つの処理ユニットに統合されてもよく、又は各ユニットは1つの独立したユニットとして使用されてもよく、2つ以上のユニットを1つのユニットに統合してもよく、上述の統合されたユニットは、ハードウェアの形で実現されてもよく、ハードウェアとソフトウェア機能ユニットの形で実現されてもよい。

【0241】

当業者は、上記した方法の実施例の全て又は一部のステップは、プログラム命令に関連するハードウェアによって完了することができ、前記プログラムは、コンピュータ読み取り可能な記憶媒体に記憶されることができ、前記プログラムが実行されるとき、上記の方法の実施例のステップを実行し、前記記憶媒体は、リムーバブルストレージ、ROM、RAM、磁気メモリ又は光ディスクなどのプログラムコードを記憶することができる媒体を含む。

40

【0242】

又は、本発明の上記の統合されたユニットがソフトウェア機能モジュールの形で実現され、スタンドアロン製品として販売又は使用される場合、コンピュータ読み取り可能な記憶媒体に記憶されてもよい。このような理解に基づいて、本発明の実施例の技術的解決策

50

は、本質的に、又は既存の技術に貢献する部分は、ソフトウェア製品の形で具現されることができ、当該コンピュータソフトウェア製品は、1つの記憶媒体に記憶されて、一台のコンピュータ機器（パーソナルコンピュータ、サーバ、又はネットワーク機器などであり得る）が本発明の各実施例に記載の方法の全部又は一部を実行するようにするためのいくつかの命令を含む。前述した記憶媒体は、リムーバブルストレージ、ROM、RAM、磁気メモリ又は光ディスクなどのプログラムコードを記憶することができる様々な媒体を含む。

【0243】

以上は、本発明の具体的な実施形態に過ぎないが、本発明の保護範囲はこれに限定されず、当業者は、本発明に開示された技術的範囲内で容易に想到し得る変更又は置換は、すべて本発明の保護範囲内に含まれるべきである。したがって、本発明の保護範囲は、特許請求の保護範囲に従うものとする。

10

20

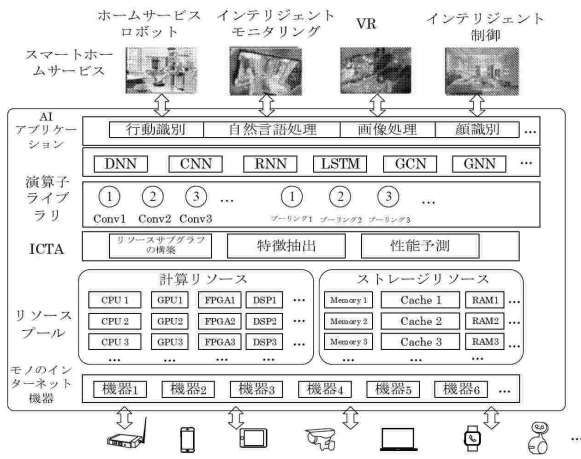
30

40

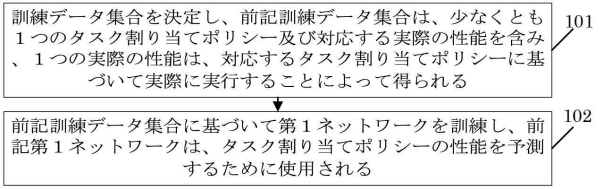
50

【 図面 】

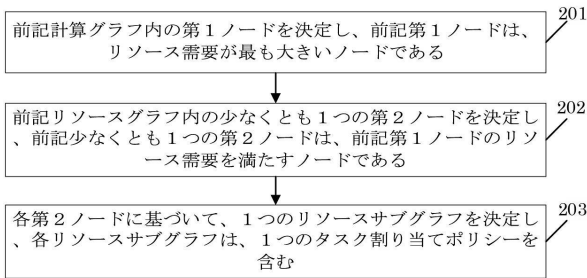
【 図 1 】



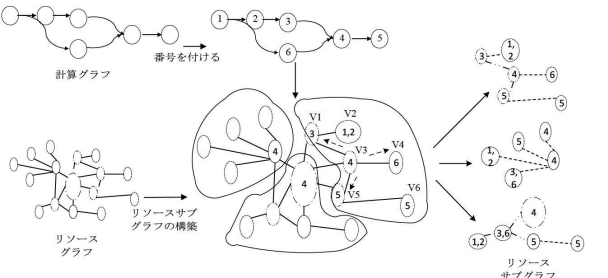
【 図 2 】



【 図 3 】



【 図 4 】



10

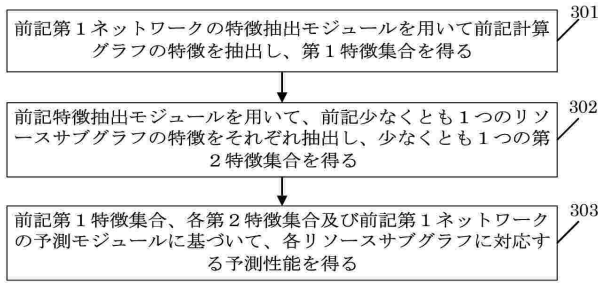
20

30

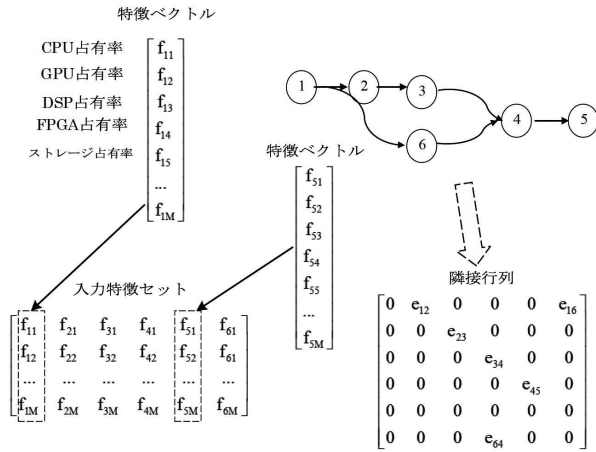
40

50

【図5】

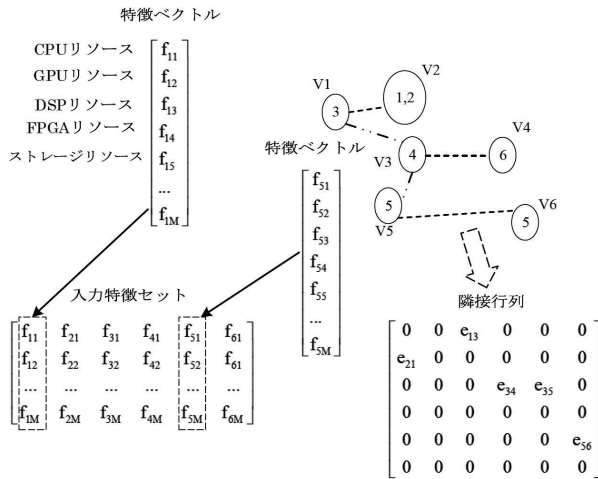


【図6a】

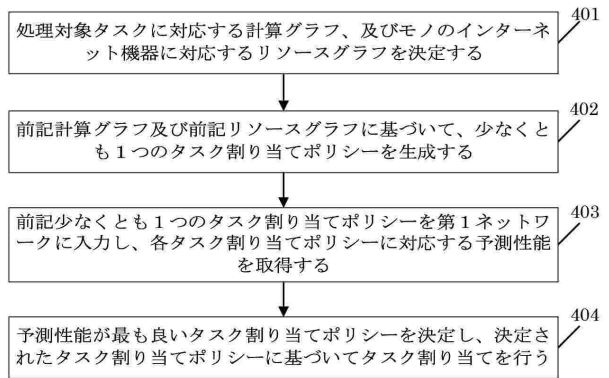


10

【図6b】



【図7】



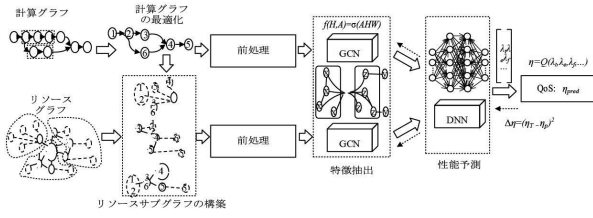
20

30

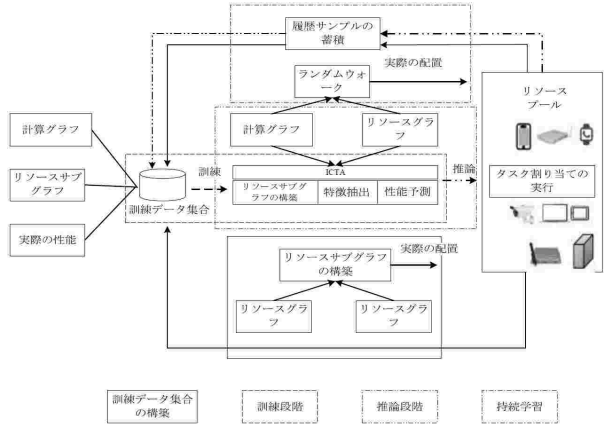
40

50

【図8】

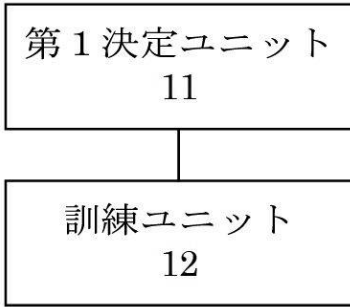


【図9】



10

【図10】



【図11】



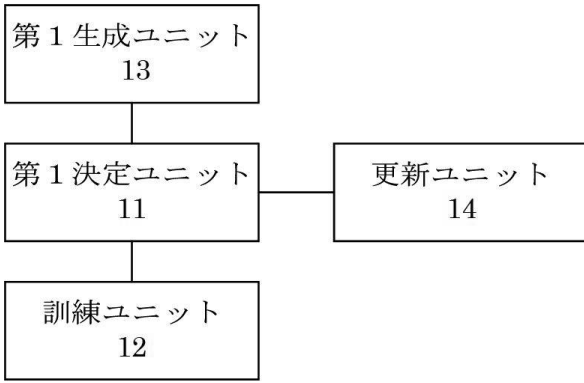
20

30

40

50

【図 1 2】



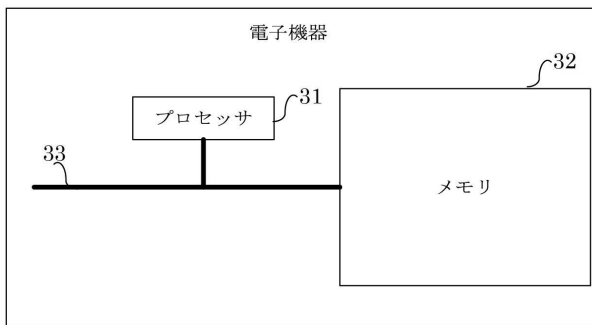
【図 1 3】



10

20

【図 1 4】



30

40

50

フロントページの続き

L T D .

中華人民共和国北京市西城区金融大街 2 9 号

(74)代理人 100078282

弁理士 山本 秀策

(74)代理人 100113413

弁理士 森下 夏樹

(74)代理人 100181674

弁理士 飯田 貴敏

(74)代理人 100181641

弁理士 石川 大輔

(74)代理人 230113332

弁理士 山本 健策

(72)発明者 曲薇

中国北京市西城区宣武 門 西大街 3 2 号

審査官 田中 幸雄

(56)参考文献 QU, Wei et al. , ICTA: Intelligent Computing Task Allocation for Efficient Deep Learning in Distributed Edge Computing System of IoT , Proceedings of the 2020 2nd International Conference on Big-data Service and Intelligent Computation , 2021年02月01日 , Pages61-69 , <https://doi.org/10.1145/3440054.3440065>

(58)調査した分野 (Int.Cl. , D B 名)

G 0 6 F 9 / 5 0