

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2019-148876
(P2019-148876A)

(43) 公開日 令和1年9月5日(2019.9.5)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 9/50 (2006.01)	G06F 9/50 150D	
G06N 20/00 (2019.01)	G06F 9/50 150C	
	G06N 99/00 150	

審査請求 未請求 請求項の数 8 O L (全 19 頁)

(21) 出願番号	特願2018-31869 (P2018-31869)	(71) 出願人	000004226 日本電信電話株式会社 東京都千代田区大手町一丁目5番1号
(22) 出願日	平成30年2月26日 (2018.2.26)	(71) 出願人	504132272 国立大学法人京都大学 京都府京都市左京区吉田本町36番地1
		(74) 代理人	100107766 弁理士 伊東 忠重
		(74) 代理人	100070150 弁理士 伊東 忠彦
		(74) 代理人	100124844 弁理士 石原 隆治
		(72) 発明者	田尻 兼悟 東京都千代田区大手町一丁目5番1号 日 本電信電話株式会社内

最終頁に続く

(54) 【発明の名称】 制御装置、データ処理制御方法、及びプログラム

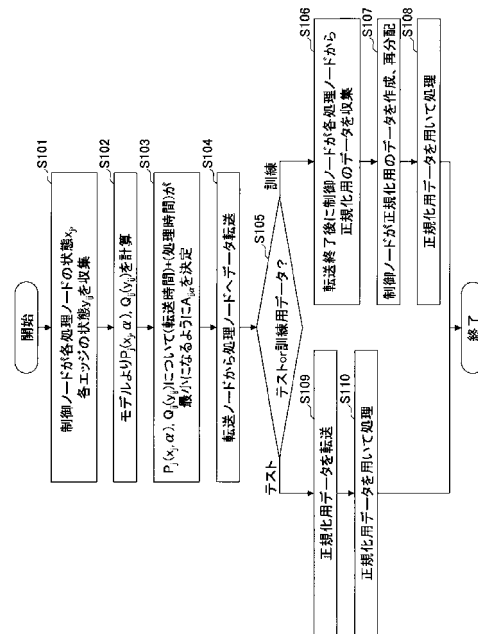
(57) 【要約】 (修正有)

【課題】複数ノードで構成されるネットワークにおいて、要求される処理精度や処理速度に応じて、データに対する処理を適切に複数ノードに分散させる。

【解決手段】データ転送機能を有するノードである転送ノードと、データ処理機能を有するノードである処理ノードと、制御機能を有するノードである制御ノードを有し、機械学習による処理を実行するネットワークにおける前記制御ノードとして機能する制御装置において、各処理ノードの負荷情報に基づいて、各転送ノードの各処理ノードへのデータの転送レートを決定する第1決定手段と、データのグルーピングと処理ノードのグルーピングに基づいて、データを処理する処理ノードを決定する第2決定手段とを備える。

【選択図】 図4

実施例5における動作例を説明するためのフローチャート



【特許請求の範囲】**【請求項 1】**

データ転送機能を有するノードである転送ノードと、データ処理機能を有するノードである処理ノードと、制御機能を有するノードである制御ノードを有し、機械学習による処理を実行するネットワークにおける前記制御ノードとして機能する制御装置であって、

各処理ノードの負荷情報に基づいて、各転送ノードの各処理ノードへのデータの転送レートを決定する第 1 決定手段と、

データのグルーピングと処理ノードのグルーピングに基づいて、データを処理する処理ノードを決定する第 2 決定手段と

を備えることを特徴とする制御装置。

10

【請求項 2】

前記第 1 決定手段は、各処理ノードの負荷情報と、転送ノードと処理ノード間の負荷情報とに基づいて、機械学習のためのデータ整形前の生データ転送時間と、処理ノードで実行されるデータ整形の時間との和が最小になるように、前記転送レートを決定する

ことを特徴とする請求項 1 に記載の制御装置。

【請求項 3】

前記第 1 決定手段は、前記負荷情報として、機械学習による回帰モデルで予測される負荷情報を使用する

ことを特徴とする請求項 2 に記載の制御装置。

【請求項 4】

前記第 2 決定手段は、処理ノード間の負荷情報と、各処理ノードでの機械学習の処理速度とに基づいて、機械学習の精度が所定の条件を満たすように、機械学習のための整形済みデータの転送時間と機械学習の処理にかかる時間との和が最小になるように、データを処理する処理ノードを決定する

ことを特徴とする請求項 1 ないし 3 のうちいずれか 1 項に記載の制御装置。

20

【請求項 5】

前記第 2 決定手段は、処理ノード間の負荷情報と、各処理ノードでの機械学習の処理速度とに基づいて、機械学習のための整形済みデータの転送時間と機械学習の処理にかかる時間との和が所定の条件を満たすように、機械学習の精度が最大になるように、データを処理する処理ノードを決定する

ことを特徴とする請求項 1 ないし 3 のうちいずれか 1 項に記載の制御装置。

30

【請求項 6】

前記第 1 決定手段は、データ転送量に関する制約条件を満足するように前記転送レートを決定する

ことを特徴とする請求項 1 ないし 5 のうちいずれか 1 項に記載の制御装置。

【請求項 7】

データ転送機能を有するノードである転送ノードと、データ処理機能を有するノードである処理ノードと、制御機能を有するノードである制御ノードを有し、機械学習による処理を実行するネットワークにおける前記制御ノードとして機能する制御装置が実行するデータ処理制御方法であって

各処理ノードの負荷情報に基づいて、各転送ノードの各処理ノードへのデータの転送レートを決定する第 1 決定ステップと、

データのグルーピングと処理ノードのグルーピングに基づいて、データを処理する処理ノードを決定する第 2 決定ステップと

を備えることを特徴とするデータ処理制御方法。

40

【請求項 8】

コンピュータを、請求項 1 ないし 6 のうちいずれか 1 項に記載の制御装置における各手段として機能させるためのプログラム。

【発明の詳細な説明】**【技術分野】**

50

【 0 0 0 1 】

本発明は、ネットワーク内外から発生するデータを分析する技術分野に関連するものである。

【 背景技術 】

【 0 0 0 2 】

IoTにおいては様々な物に組み込まれたセンサから得られる情報がネットワークを通じて通信される。それらを収集し分析し、意味のある情報を抽出することによって新たなアプリケーションの実現が可能となる（非特許文献1）。また一方でIoTにおいてセンサが接続するネットワーク自体もログデータ等多数の情報を出力しており、これを分析することも新たなアプリケーションが実現しうる。

10

【 0 0 0 3 】

しかし、多数のセンサがネットワークに接続した場合、それに伴いネットワーク上に膨大なセンサデータが存在することになり、ネットワークの負荷及び膨大なデータを分析するサーバの負荷が課題となる。このような問題を解決するために複数のサーバで分散して処理を行うことで負荷を軽減しようとする研究がなされている。例えば非特許文献2では画像分析において画像からテキストへの前処理をエッジサーバとクラウドサーバで分散する手法を提案している。

【 先行技術文献 】

【 非特許文献 】

【 0 0 0 4 】

20

【 非特許文献 1 】 Charith Perera, Arkady Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos. "Context aware computing for the internet of things: A survey," Communications Surveys & Tutorials, IEEE, Vol. 16, No. 1, pp. 414-454, 2014.

【 非特許文献 2 】 新熊亮一・加藤慎悟・上林将大・池田泰弘・川原亮一・林 孝典 " エッジ-クラウド連携制御のためのシステム設計の研究", 信学技報, vol. 117, no. 71, MoNA2017-2, pp. 29-32, 2017年6月.

【 発明の概要 】

【 発明が解決しようとする課題 】

【 0 0 0 5 】

30

しかし、上記の研究では機械学習のテストにおける前処理の負荷分散を議論したのみであり、IoTにおける包括的な負荷軽減を考えるならば、データの前処理、訓練、テストの全ての段階において動的に負荷分散がなされるべきであると考えられる。また、ネットワーク自身も分析対象となるデータを生み出す点において、自分自身も分析対象として扱い、それに応じてオペレーションされることにより強固なIoT基盤が実現すると考えられる。

【 0 0 0 6 】

本発明は上記の点に鑑みてなされたものである、複数ノードで構成されるネットワークにおいて、要求される処理精度や処理速度に応じて、データに対する処理を適切に複数ノードに分散させることを可能とする技術を提供することを目的とする。

【 課題を解決するための手段 】

40

【 0 0 0 7 】

開示の技術によれば、データ転送機能を有するノードである転送ノードと、データ処理機能を有するノードである処理ノードと、制御機能を有するノードである制御ノードを有し、機械学習による処理を実行するネットワークにおける前記制御ノードとして機能する制御装置であって、

各処理ノードの負荷情報に基づいて、各転送ノードの各処理ノードへのデータの転送レートを決定する第1決定手段と、

データのグルーピングと処理ノードのグルーピングに基づいて、データを処理する処理ノードを決定する第2決定手段と

を備えることを特徴とする制御装置が提供される。

50

【発明の効果】

【0008】

開示の技術によれば、複数ノードで構成されるネットワークにおいて、要求される処理精度や処理速度に応じて、データに対する処理を適切に複数ノードに分散させることを可能とする技術が提供される。

【図面の簡単な説明】

【0009】

【図1】本発明の実施の形態におけるシステム構成の例を示す図である。

【図2】制御装置の機能構成の一例を示す図である。

【図3】制御装置のハードウェア構成の一例を示す図である。

10

【図4】実施例5における動作例を説明するためのフローチャートである。

【図5】実施例6における動作例を説明するためのフローチャートである。

【図6】実施例7における動作例を説明するためのフローチャートである。

【図7】実施例8における動作例を説明するためのフローチャートである。

【図8】効果を説明するためのラダー型ネットワークの構成例を示す図である。

【図9】クラウドエッジ対応による処理速度と精度の関係性を示す図である。

【発明を実施するための形態】

【0010】

以下、図面を参照して本発明の実施の形態（本実施の形態）を説明する。以下で説明する実施の形態は一例に過ぎず、本発明が適用される実施の形態は、以下の実施の形態に限られるわけではない。

20

【0011】

以下では、本実施の形態におけるシステム（ネットワーク）構成例を実施例1として説明し、当該システムにおける動作例を実施例2～9において説明する。なお、実施例1～9は任意に組み合わせて実施することが可能である。

【0012】

（実施例1）

<全体構成>

本実施の形態におけるシステム（ネットワーク）は、ネットワークの境界に存在するエッジノードと、ネットワーク内部に存在するコアノードから構成される。

30

【0013】

当該ネットワークの外部にはセンサデバイスが存在し、エッジノードが当該センサデバイスにより生成されるセンサデータを収集する。また、各エッジノード及び各コアノードのそれぞれからログデータが得られる。本実施の形態では、センサデータ及びログデータの2種類のデータを取り扱う。各ノードは、例えば、データの送受信機能を有するコンピュータ（ルータ等を含む）である。また、当該コンピュータは、仮想マシン（クラウドで提供されるサーバや通信装置等）であってもよい。

【0014】

ネットワークを構成する各ノードは、データ転送機能を持つノード（以下、転送ノードと称する）、データ処理機能を持つノード（以下、処理ノードと称する）、制御機能を持つノード（以下、制御ノードと称する）の3種類に分類される。ただし、1つのノードが、複数の機能を備えてもよい。図1は、本実施の形態におけるネットワークの構成の一例を示している。

40

【0015】

転送ノードは処理ノードにデータを転送する。処理ノードは、当該データに対し、機械学習による分析を行うための数値ベクトル化、正規化等のデータ整形（前処理と称してもよい）を行う。制御ノードは、どの処理ノードにどのデータを処理させるかと、転送ノードがどのノードにどれくらいデータを送るのかを決定する。

【0016】

より具体的には、外部センサからエッジノードに入力される外部センサデータとしては

50

、例えば、カメラ画像、各センサの感知した温度、圧力等センサに応じた様々なデータがある。

【 0 0 1 7 】

また、ネットワーク内部から生成されるログデータとしては、例えば、NetFlowデータ、MIBデータ、Syslogデータ等がある。

【 0 0 1 8 】

ネットワークにおけるどのノードを、制御ノード、処理ノード、転送ノードのどれにするかについては、例えば、ネットワークトポロジーにおける位置関係やハードの処理能力等に従って決定する。前述したように、1つのノードが複数の機能を持つことも可能である。この決定はハードの交換等がない限りは固定とするが、固定であることに限定されるわけではなく、ノードの機能が変更される場合があってもよい。

10

【 0 0 1 9 】

< 制御ノードの機能構成 >

図2に、制御ノードの構成例を示す。制御ノードを「制御装置」と称してもよい。図2に示すように、本実施の形態における制御ノードは、入力部101、記憶部102、モデル計算部103、正規化パラメータ管理部104、機械学習テスト部105、転送比率計算部106、機械学習訓練部107、出力部108を有する。図2に示す構成は、制御ノードが、処理ノードの機能も含む例を示している。各機能部の動作概要は次のとおりである。

【 0 0 2 0 】

入力部101は、各ノードから、処理対象データや、制御に用いる情報（負荷情報等）を入力する。記憶部102は、予め作成したモデルのパラメータや、処理に使用する各種データを格納する。

20

【 0 0 2 1 】

モデル計算部103は、モデルを用いた計算を実行する正規化パラメータ管理部104は、実施例2、5において説明する正規化に係る処理を実行する。機械学習テスト部105は、機械学習の訓練で得られたモデルを用いたテストを行う。転送比率計算部106は、実施例2、3等において説明する転送レートに係る処理を実行する。機械学習訓練部107は、機械学習の訓練を実施し、モデル（パラメータ等）を決定する。出力部108は、正規化パラメータ管理部104や転送比率計算部106等の計算により得られた結果を対象ノードへ送信する等の動作を行う。

30

【 0 0 2 2 】

< ハードウェア構成例 >

上述した制御ノード（制御装置）は、例えば、コンピュータに、本実施の形態で説明する処理内容を記述したプログラムを実行させることにより実現可能である。

【 0 0 2 3 】

すなわち、制御ノードは、コンピュータに内蔵されるCPUやメモリ等のハードウェア資源を用いて、制御ノードで実施される処理に対応するプログラムを実行することによって実現することが可能である。上記プログラムは、コンピュータが読み取り可能な記録媒体（可搬メモリ等）に記録して、保存したり、配布したりすることが可能である。また、上記プログラムをインターネットや電子メール等、ネットワークを通して提供することも可能である。

40

【 0 0 2 4 】

図3は、本実施の形態における上記コンピュータのハードウェア構成例を示す図である。図3のコンピュータは、それぞれバスBで相互に接続されているドライブ装置150、補助記憶装置152、メモリ装置153、CPU154、インターフェース装置155、表示装置156、及び入力装置157等を有する。

【 0 0 2 5 】

当該コンピュータでの処理を実現するプログラムは、例えば、CD-ROM又はメモリカード等の記録媒体151によって提供される。プログラムを記憶した記録媒体151が

50

ドライブ装置 150 にセットされると、プログラムが記録媒体 151 からドライブ装置 150 を介して補助記憶装置 152 にインストールされる。但し、プログラムのインストールは必ずしも記録媒体 151 より行う必要はなく、ネットワークを介して他のコンピュータよりダウンロードするようにしてもよい。補助記憶装置 152 は、インストールされたプログラムを格納すると共に、必要なファイルやデータ等を格納する。

【0026】

メモリ装置 153 は、プログラムの起動指示があった場合に、補助記憶装置 152 からプログラムを読み出して格納する。CPU 154 は、メモリ装置 153 に格納されたプログラムに従って、制御ノードに係る機能を実現する。インターフェース装置 155 は、ネットワークに接続するためのインターフェースとして用いられる。図 2 に示した機能構成における入力部 101 と出力部 108 は、当該インターフェース装置 155 により実現することが可能である。

10

【0027】

表示装置 156 はプログラムによる GUI (Graphical User Interface) 等を表示する。入力装置 157 はキーボード及びマウス、ボタン、又はタッチパネル等で構成され、様々な操作指示を入力させるために用いられる。なお、制御ノードにおいて入力装置 157 と表示装置 156 を備えないこととしてもよい。

【0028】

(実施例 2)

実施例 2 では、実施例 1 で説明した、各機能が各ノードに配置されたネットワークにおいて、制御ノードの転送比率計算部 106 が、各転送ノードで収集されるログデータから得られる現在の負荷状況(CPU使用率、転送量等)等に基づき、データ処理や転送による遅延、転送量の観点から、各転送ノードがどの処理ノードにどのくらいの比率でどのデータを転送するかを制御する。

20

【0029】

また、制御ノードの正規化パラメータ管理部 104 は、機械学習において必要となるデータの正規化を行うために、同じ種類のデータが複数の処理ノードに分散した場合において、各処理ノードの持っているデータのうち正規化に必要な情報を集約計算し、再分配する。

【0030】

より詳細には、制御ノードの入力部 101 を介して転送比率計算部 106 が収集する負荷状況の具体例としては、例えば、各ノードの CPU 使用率や、各リンクのスループット、RTT 等がある。

30

【0031】

また、制御ノードの転送比率計算部 106 が情報を収集する方法としては、例えば、SDN 等、ノード及びエッジの負荷状況を直接観測できる基盤において直接収集する方法、top コマンドや ping コマンドで CPU 使用率、ネットワーク遅延等を測定する方法、NetFlow や MIB 等のログデータから抽出する方法等がある。なお、「エッジ」とは、ノード間のネットワーク部分を示す。

【0032】

各ノードや各エッジの負荷状況、及びデータ整形方法に基づく転送速度やデータ整形速度は予めモデル化しておき、制御ノードの転送比率計算部 106 は記憶部 102 に当該モデルを保持し、当該モデル及び現在の負荷状況(CPU使用率、転送量等)に基づき、データ処理や転送による遅延、転送量の観点から、各転送ノードがどの処理ノードにどのくらいの比率でどのデータを転送するかを制御する。転送比率計算部 106 は、データ転送及び整形にかかる時間が最短になるようにデータの転送制御を行う。

40

【0033】

ここで、機械学習における訓練時の各データの正規化について、例えば $(value - \text{最小値}) / (\text{最大値} - \text{最小値})$ のように行くとすると、制御ノードの正規化パラメータ管理部 104 は、各処理ノードが持つデータの最小値 \min_j 及び最大値 \max_j を入力部 101 を介して収集

50

し、これから全体の最小値= $\min\{\min_j\}$, 最大値= $\max\{\max_j\}$ を計算し、当該計算された最小値と最大値を当該データを持つ全処理ノードに再分配する。

【0034】

一方、テスト時においては、制御ノードの正規化パラメータ管理部104は、訓練の際に決定した最小値と最大値を各処理ノードに転送し、各処理ノードはこれを用いて処理を行う。

【0035】

上述したように、制御ノードは、データ転送及び整形にかかる時間が最短になるようにデータの転送制御を行うが、転送ノードに存在するデータはテスト用に処理ノードに転送された際にも転送ノードにも保持されており、これはテスト終了後に訓練用のデータとして再度用いられる。ただし、機械学習が異常検知である場合は、テストに異常と判断されたデータは学習には用いられない。

10

【0036】

実施例2におけるデータ転送制御に係るより具体的な処理内容については実施例5において説明する。

【0037】

(実施例3)

実施例3では、実施例1におけるネットワーク形状や対象データを前提として、制御ノードの転送比率計算部106が、処理ノードから得られる、データ転送時間を含むデータ分析時間、及びデータ分析の精度に基づいて、各転送ノードから得られたデータをグルーピングして機械学習のモデルパラメータを決定し、更に分散学習のために処理ノードをグルーピングし、どの処理ノードグループで訓練を行うかを決定し、当該処理ノードグループにおいて訓練を実施させる。なお、単一処理ノードで訓練を行うことも可能である。

20

【0038】

以上の訓練結果を利用して、制御ノードの転送比率計算部106は、テストを実行する処理ノードグループを決定し、当該処理ノードグループの処理ノードにテストを行わせる。この際テストを行う処理ノードグループは訓練の時に用いたものと違ってよい。

【0039】

より具体的には、初めに、訓練やテストに使用する処理ノードグループ、データのグループ化方法及び学習モデルのパラメータを与えるとどのくらいの処理速度及び精度が出るのかを予めモデル化し、当該モデルを記憶部102に格納しておく。そして、転送比率計算部106は、当該モデルに従って処理速度、機械学習の精度が与えられた条件を満たすように機械学習のモデルパラメータを制御していく。なお、この制御の中で、モデルを用いた計算は、モデル計算部103により行われる。制御の中で、モデルとの乖離が現れた場合、各種パラメータと現在の状態を使ってモデルを更新してもよい。

30

【0040】

より詳細な処理ノードの決定方法、データグルーピングの決定方法等については実施例6、7において説明する。

【0041】

(実施例4)

実施例2においてログデータからノードやエッジの負荷状況を抽出する場合、ログデータを実際に処理する時間が存在し、結果として制御ノードが得るネットワークの負荷状況と現在時刻に時間差が生じることになる。そこで、実施例4では、ログデータを実施例3に基づいて分析し予測を行い、その結果と実施例3で説明したモデルを用いて、制御ノードの転送比率計算部106がデータ転送制御を行う。予測の方法については例えば回帰分析等を用いる。

40

【0042】

制御方法の詳細については、実施例8で説明する。

【0043】

(実施例5)

50

実施例 5 では、実施例 2 の詳細例を説明する。実施例 5 において、各転送ノードには単位時間当たりそれぞれの種類のデータが N_i だけ入力されるとする。また、訓練に用いるデータの期間を T_s 、テストに用いるデータの期間を T_p とする。

【 0 0 4 4 】

また、処理ノード j の状態 (CPU 使用率等) を x_j とし、ノード ij を接続するエッジの状態 (ネットワーク負荷状況等) を y_{ij} とし、これらについて処理ノードの前処理速度 P_j 及びエッジの転送速度 Q_{ij} が x_j, y_{ij} の関数として $P_j(x_j, y_{ij})$ [bps] (ここで y_{ij} はデータの種類を表す。種類としては、NetFlow のデータ、MIB のデータ、センサから得られる温度情報等がある)、 $Q_{ij}(y_{ij})$ [bps] のようにモデリングされているとする ($i = 1 \sim n$: 転送ノード, $j = 1 \sim m$: 処理ノード, $i = j$ は同じノードを表すので $Q_{jj} = \text{inf.}$ としておく)。ノード ij は、処理ノード j にデータを転送する転送ノード i である。

10

【 0 0 4 5 】

上記モデル (パラメータ等) は、制御ノードの記憶部 102 に格納される。また、データ整形の際、各種類のデータの整形は同時に行うことが可能であるとする。

【 0 0 4 6 】

実施例 5 における処理手順を図 4 のフローチャートを参照して説明する。S101 において、制御ノードの転送比率計算部 106 は、各処理ノードの状態 x_j 、及び、各エッジの状態 y_{ij} を収集する。S102 において、転送比率計算部 106 は、上述したモデルに基づいて、 $P_j(x_j, y_{ij})$ と $Q_{ij}(y_{ij})$ を計算する。

【 0 0 4 7 】

S103 において、転送比率計算部 106 は、 $P_j(x_j, y_{ij})$ と $Q_{ij}(y_{ij})$ に基づいて、ネットワーク全体での (生データ転送時間) + (データ整形時間) が最小になるように、各転送ノードの各処理ノードへのデータ転送レートであるデータ整形前転送レート A_{ij} を決定する。

20

【 0 0 4 8 】

A_{ij} の決定方法を説明する。訓練の場合はデータの転送が完了した後、処理ノードと制御ノードが連携して正規化パラメータを定めてデータ整形を行うことになるので、データの転送からデータ整形までの時間を最速にするような A_{ij} を決定する式は、

$$\text{argmin}_{\{A_{ij}\}} (\max_{\{i, j\}} (N_i \cdot T_s / Q_{ij}(y_{ij})) + \max_{\{j\}} ((\sum_{i=1}^n N_i \cdot T_p) / P_j(x_j, y_{ij})))$$

30

と表される。転送比率計算部 106 は、当該式に示される最適化問題を解くことで、訓練時の A_{ij} を決定する。

【 0 0 4 9 】

テスト時の場合、既に正規化パラメータは用意してあるのでデータの転送から整形までの時間を最速にする A_{ij} を決定する式は、

$$\text{argmin}_{\{A_{ij}\}} (\max_{\{j\}} (z_j / Q_{ij}(y_{ij})) + \max(\max_{\{i\}} (A_{ij} \cdot (N_i - z_j) \cdot T_p / Q_{ij}(y_{ij})), \sum_{i=1}^n (A_{ij} \cdot N_i \cdot T_p / P_j(x_j, y_{ij}))))$$

と表わされる。ここで 2 項目の $\max()$ の中の 1 項目はデータの転送速度が処理速度より遅い場合、2 項目はデータの処理速度が転送速度より遅い場合を表す。

【 0 0 5 0 】

図 4 の S103 において、転送比率計算部 106 は、これらの式より訓練時、テスト時それぞれの場合において A_{ij} を決定する。

40

【 0 0 5 1 】

転送比率計算部 106 で計算された A_{ij} は、出力部 108 から該当の転送ノードに通知される。S104 では、当該 A_{ij} に基づいて、転送ノードから処理ノードへのデータ転送が行われる。

【 0 0 5 2 】

図 4 の S105 ~ S110 は、実施例 2 で説明したデータの正規化に関する処理を示している。

【 0 0 5 3 】

50

S 1 0 5 における判定が「訓練」の場合、S 1 0 6 において、制御ノードの正規化パラメータ管理部 1 0 4 は、各処理ノードが持つ正規化用のデータ（例：前述したデータの最小値 \min_j 及び最大値 \max_j ）を収集する。

【 0 0 5 4 】

S 1 0 7 において、制御ノードの正規化パラメータ管理部 1 0 4 は、収集したデータに基づいて、正規化用のデータ（例：前述した最小値 $=\min\{\min_j\}$, 最大値 $=\max\{\max_j\}$ ）を作成し、当該作成された正規化用データを各処理ノードに再分配する。S 1 0 8 において、各処理ノードは、当該正規化用データを用いて機械学習等の処理を行う。

【 0 0 5 5 】

S 1 0 5 における判定が「テスト」の場合、制御ノードの正規化パラメータ管理部 1 0 4 は、訓練の際に決定した正規化用データを各処理ノードに転送し（S 1 0 9 ）、各処理ノードはこれを用いて処理を行う（S 1 1 0 ）。

【 0 0 5 6 】

（実施例 6 ）

次に、実施例 6 を説明する。実施例 6 は実施例 3 の詳細例である。実施例 6 では、制御ノードの転送比率計算部 1 0 6 が、データ処理時間や機械学習の要求精度が予め定めた条件を満たすように、データのグルーピング、及び処理ノードのグルーピングを行う。これらグルーピングに従って、データの再配分がなされる。より詳細には下記のとおりである。

【 0 0 5 7 】

ここでは、前処理を行った処理ノードから機械学習を行う処理ノードに対して転送されるデータのグルーピングを G_d とし、処理ノードのグルーピングを G_n とし、訓練に用いるデータの期間を T_s 、テストに用いるデータの期間を T_p とした上で、機械学習の精度を G_d 、 T_s 、 T_p を変数とする関数 $F(G_d, T_s, T_p)$ として定め、制御ノードの転送比率計算部 1 0 6 が、指定された分析精度 に対して $F(G_d, T_s, T_p) >$ を満たすという条件を満たすように、各処理ノードの機械学習の訓練速度 $R(x(g_{nk}), g_{dl})$ 、テスト速度 $R'(x(g_{nk}), g_{dl})$ 、及び $Q_{jj'}(y_{jj'})$ に基づき、（整形済みデータのデータ転送時間）+（訓練もしくはテストにかかる時間）を最小にするように、訓練に用いるデータの期間 T_s 、テストに用いるデータの期間 T_p 、データ整形後転送レート $B_{jj'}G$ （ j は前処理を行った処理ノードで、 j' は機械学習を行う処理ノード）、データのグルーピング G_d 、及び処理ノードのグルーピング G_n の最適化を行う。

【 0 0 5 8 】

上記の G_d はデータの種類 $_k$ について $G_d = \{g_{d1}, g_{d2}, \dots\} = \{\{ _1, _2, \dots\}, \{ _k, \dots\}, \dots\}$ のようにグルーピングを行うことを示す。なお、 $_1$ はデータの種類の1次元目を意味する。また、モデルグループの中に複数の種類のデータの列が含まれていてもよい。また、処理ノードのグルーピング G_n は、処理ノード j について $G_n = \{g_{n1}, g_{n2}, \dots\} = \{\{1, 2, \dots\}, \{j, \dots\}, \dots\}$ のようにグルーピングを行うことを示す。

【 0 0 5 9 】

また、 $x(g_{nk})$ はグループ g_{nk} に含まれるノード j の状態の集合を意味する。なお、 G_d は訓練とテストで共通だが、 G_n は訓練とテストで異なってもよくそれぞれのフェーズで最速になるように決定される。

【 0 0 6 0 】

また、 G_d の各要素をどの G_n で処理を行うかを定義する記号として $G = \{g_1, g_2, \dots\} = \{(g_{d1}, g_{na}), (g_{d2}, g_{nb}), \dots\}$ という表記を定義する。また、 $B_{jj'}G$ は、処理ノード j' が属するノードグループ内における、前処理を行った処理ノードから機械学習を行う処理ノードへの転送レートを示す。前処理を行った処理ノード j と機械学習を行う処理ノード j' は別のノードグループに属していてもよい。 $B_{jj'}G$ は、同一ノードグループ内のどのノードにどのくらいデータを送るのかということに自由度がある。

【 0 0 6 1 】

実施例 6 の処理手順例を図 5 のフローチャートを参照して説明する。

10

20

30

40

50

【 0 0 6 2 】

S 2 0 1において、制御ノードの転送比率計算部 1 0 6 は、入力部 1 0 1を介して各処理ノードの状態 x_j 、各エッジの状態 $y_{jj'}$ を収集する。

【 0 0 6 3 】

訓練を行う場合 (S 2 0 2の訓練)、S 2 0 3に進み、制御ノードの転送比率計算部 1 0 6 は、モデルに基づいて、様々なGについて訓練速度 $R(x(g_{nk}), g_{dl})$ 、及び機械学習の精度 $F(G_d, T_s, T_p)$ を計算する。

【 0 0 6 4 】

S 2 0 4において、制御ノードの転送比率計算部 1 0 6 は、 $R(x(g_{nk}), g_{dl})$ 、 $Q_{jj'}$ ($y_{jj'}$)、 $F(G_d, T_s, T_p)$ に基づいて、 $F(G_d, T_s, T_p) >$ を満たすという条件のもとで T_s が最小になるようにG、 $B_{jj'}$ G、 T_s を決定する。 10

【 0 0 6 5 】

T_s は、(整形済みデータのデータ転送時間)+(訓練にかかる時間)であり、

$$T_s = \max_j \{ B_{jj'} G^* (\sum_i (A_{ij} * N_i * T_s)) / Q_{jj'} \} + \max_G \{ G (\sum_{jj'} (B_{jj'} G^* (\sum_i (A_{ij} * N_i * T_s))) / R(x(g_{nk}), g_{dl})) \}$$

と表わされる。転送比率計算部 1 0 6 は、当該最適化問題を解くことにより、G、 $B_{jj'}$ G、 T_s を決定する。ここで、 g_{dl} 、 j' g_{nk} である。

【 0 0 6 6 】

テストを行う場合 (S 2 0 2のテスト)、S 2 0 5に進み、制御ノードの転送比率計算部 1 0 6 は、モデルに基づいて、様々なGについてテスト速度 $R'(x(g_{nk}), g_{dl})$ 、及び機械学習の精度 $F(G_d, T_s, T_p)$ を計算する。 20

【 0 0 6 7 】

S 2 0 6において、制御ノードの転送比率計算部 1 0 6 は、 $R'(x(g_{nk}), g_{dl})$ 、 $Q_{jj'}$ ($y_{jj'}$)、 $F(G_d, T_s, T_p)$ に基づいて、 $F(G_d, T_s, T_p) >$ を満たすという条件のもとで T_p が最小になるように T_p 、 $B_{jj'}$ Gを決定する。

【 0 0 6 8 】

T_p は、(整形済みデータのデータ転送時間)+(テストにかかる時間)であり、

$$T_p = \max_j \{ B_{jj'} G^* (\sum_i (A_{ij} * N_i * T_p)) / Q_{jj'} \} + \max_G \{ G (\sum_{jj'} (B_{jj'} G^* (\sum_i (A_{ij} * N_i * T_p))) / R'(x(g_{nk}), g_{dl})) \}$$

と表わされる。転送比率計算部 1 0 6 は、当該最適化問題を解くことにより、 T_p 、 $B_{jj'}$ Gを決定する。 30

【 0 0 6 9 】

上記の計算により得られた結果は、出力部 1 0 8から対象の各ノードに通知され、各ノードは、計算で得られたパラメータに基づいて処理を行う。

【 0 0 7 0 】

(実施例 7)

次に、実施例 7を説明する。実施例 7も実施例 3の詳細例である。実施例 7でも制御ノードの転送比率計算部 1 0 6 が、データ処理時間や機械学習の要求精度が予め定めた条件を満たすように、データのグルーピング、及び処理ノードのグルーピングを行う。より詳細には下記のとおりである。実施例 7は、以下では、主に、実施例 6と異なる点を説明する。 40

【 0 0 7 1 】

実施例 6で説明したとおり、処理ノードに前処理後のデータが配置されている場合、((機械学習に用いる処理ノードに転送する時間)+(機械学習処理にかかる時間))は訓練、テストそれぞれで T_s 、 T_p である。

実施例 7では、訓練及びテストを行う時間間隔をそれぞれ T_s 、 T_p とする。 T_s 及び T_p はシステムの要件として定められるパラメータである。

【 0 0 7 2 】

そして、転送比率計算部 1 0 6 は、 $\min \{ T_s, (\text{生データの転送時間}) + (\text{データ整形時間}) + (\text{整形済みデータのデータ転送時間}) + (\text{訓練にかかる時間}) \} <$ T_s を満たすという条件のもと 50

とで $F(G_d, T_s, T_p)$ を最大化するように $G, B_{jj}'G, T_s$ を決定する。また、転送比率計算部106は、 $\min\{T_p, (\text{生データの転送時間})+(\text{データ整形時間})+(\text{整形済みデータのデータ転送時間})+(\text{テストにかかる時間})\} < T_p$ を満たすという条件のもとで $F(G_d, T_s, T_p)$ を最大化するように $G, B_{jj}'G, T_p$ を決定する。

【0073】

言い換えると、転送比率計算部106は、システム要件として要求される T_s, T_p に対して $\max(T_s, T_s) < T_s, \max(T_p, T_p) < T_p$ を満たすという条件のもとで機械学習の精度 $F(G_d, T_s, T_p)$ を最大化するように、 $G, T_s, T_p, B_{jj}'G$ を決定する。

【0074】

実施例7の処理手順例を図6のフローチャートを参照して説明する。

10

【0075】

S301において、制御ノードの転送比率計算部106は、入力部101を介して各処理ノードの状態 x_j 、各エッジの状態 y_{jj}' を収集する。

【0076】

訓練を行う場合(S302の訓練)、S303に進み、制御ノードのモデル計算部103は、モデルに基づいて、様々な G について訓練速度 $R(x(g_{nk}), g_{dl})$ 、及び機械学習の精度 $F(G_d, T_s, T_p)$ を計算する。

【0077】

S304において、制御ノードのモデル計算部103は、 $R(x(g_{nk}), g_{dl}), Q_{jj}'(y_{jj}')$ 、 $F(G_d, T_s, T_p)$ に基づいて、 $\max\{T_s, T_s\} < T_s$ を満たすという条件のもとで、 $F(G_d, T_s, T_p)$ が最大になるように $G, B_{jj}'G, T_s$ を決定する。

20

【0078】

テストを行う場合(S302のテスト)、S305に進み、制御ノードのモデル計算部103は、モデルに基づいて、様々な G についてテスト速度 $R'(x(g_{nk}), g_{dl})$ 、及び機械学習の精度 $F(G_d, T_s, T_p)$ を計算する。

【0079】

S306において、制御ノードの転送比率計算部106は、 $R'(x(g_{nk}), g_{dl}), Q_{jj}'(y_{jj}')$ 、 $F(G_d, T_s, T_p)$ に基づいて、 $\max\{T_p, T_p\} < T_p$ を満たすという条件のもとで、 $F(G_d, T_s, T_p)$ が最大になるように $T_p, B_{jj}'G$ を決定する。

30

【0080】

上記の計算により得られた結果は、出力部108から対象の各ノードに通知され、各ノードは、計算で得られたパラメータに基づいて処理を行う。

【0081】

(実施例8)

次に、実施例8を説明する。実施例8は実施例4の詳細例である。実施例8では、処理ノードにおける機械学習として回帰分析が行われているとする。周期 T_s, T_p で各ノードの状態 x_j 及び各エッジの状態 y_{ij} や y_{jj}' に対する回帰の訓練とテストを行っている場合において、制御ノードは、ある周期の訓練やテストが始まる前の1周期前に行われたテストの結果で得られた回帰予測値 $x_{pj}(t - T_p)$ (各処理ノードの状態)、 $y_{p_{ij}}(t - T_p)$ 、 $y_{p_{jj}'}(t - T_p)$ (各エッジの状態)に基づいて A_{ij} 、 $G, B_{jj}'G, T_s, T_p$ を制御する。ここで、 $x_{pi}(t - T_p)$ は、ノードにおける時刻 $t - T_p$ の状態に基づく時刻 t についての予測値を表し、 $y_{p_{ij}}(t - T_p)$ 及び $y_{p_{jj}'}(t - T_p)$ はエッジにおける時刻 $t - T_p$ の状態に基づく時刻 t についての予測値を表す。

40

【0082】

すなわち、実施例8では、制御ノードは、現在時刻 t から時刻 T_p だけ前の処理ノードの状態 $x_{pj}(t - T_p)$ 及び各エッジの状態 $y_{p_{ij}}(t - T_p)$ 、 $y_{p_{jj}'}(t - T_p)$ に対して、転送レート A_{ij} 及び $B_{jj}'G$ が与えられた時の $x_{pj}(t), y_{p_{ij}}(t)$ 及び $y_{p_{jj}'}(t)$ 、それらに基づく $P_j(x_{pj}(t), \quad), Q_{ij}(y_{p_{ij}}(t)), Q_{jj}'(y_{p_{jj}'}(t)), R(x(g_{nk}(t)), g_{dl}), R'(x(g_{nk}(t)), g_{dl})$ を実施例2、3等の手順に従って計算し、データ整形及び転送にかかる時間を最速にする A_{ij} 及び $B_{jj}'G$ を決定する。

50

【0083】

図7のフローチャートを参照して、実施例8の処理手順を説明する。S401において、制御ノードの転送比率計算部106は、時刻 t において時刻 $t-p$ で行った回帰予測の結果 $x_{p_j}(t-p)$, $y_{p_{ij}}(t-p)$ を用いて実施例2の処理を実施する。すなわち、転送比率計算部106は、 A_{ij} を決定して、処理ノードに前処理を行わせる。

【0084】

S402において、制御ノードの転送比率計算部106は、時刻 t において時刻 $t-p$ で行った回帰予測の結果 $x_{p_j}(t-p)$, $y_{p_{jj}}'(t-p)$ を用いて実施例3の処理を実施し、回帰予測により $x_{p_j}(t)$, $y_{p_{jj}}'(t)$ を予測する。

【0085】

S403において、制御ノードの転送比率計算部106は、 t を $t+P$ に更新し、S401からの処理を繰り返す。

【0086】

(実施例9)

次に、実施例9を説明する。実施例9では、実施例5、6、7において、制御ノードの転送比率計算部106が、データ転送レート A_{ij} , $B_{jj}'G$ を決定する際に、データ転送量に関する制約条件 $H(A_{ij}, B_{jj}'G)$ を満足するような A_{ij} 及び $B_{jj}'G$ を決定する。なお、制約条件は、 $H_1(A_{ij})$ 1、 $H_2(B_{jj}'G)$ 2のように、 A_{ij} , $B_{jj}'G$ のそれぞれについて表わされてもよい。

【0087】

ここで、 $H(A_{ij}, B_{jj}'G)$ は、例えば、転送ノード(あるいは処理ノード)が処理ノードにデータを転送することによる転送量を表す関数であり、一例としては、 A_{ij} 及び $B_{jj}'G$ の全ての値についての統計量(総和, 平均, 最大値等)である。また、当該関数として、 A_{ij} や $B_{jj}'G$ の値だけではなく、ノード i からノード j までのホップ数も掛け合わせる等、ネットワークに対する負荷を考慮した関数を使用することとしてもよい。

【0088】

(効果について)

実施例1~9を用いて説明した本発明に係る技術により、ネットワーク内外から発生する膨大な情報を要求される精度及び速度で分析を行う際に、サンプリングレート及び分析と処理を行うノードを動的に決定し、負荷分散及び要求される精度を実現するモデリングを行うことができる。すなわち、当該技術により、複数ノードでネットワークが構成される系において、要求される精度や処理速度に応じて機械学習を分散させて行うことが可能となる。ここでは、実施例7に従い、データを分散して処理を行う場合に求められる処理速度を保持したまま、どれだけの精度が確保できるのかについて説明する。

【0089】

ここでは、一例として、図8に示すネットワーク構造を対象とし、実施例2、5で説明した前処理は全てのノードで自身の保有するデータについて行ったものとし、各ノードから発生するMIBデータとSyslogデータを用いた異常検知を行うものとする。また、各ノードで発生したこれらのデータはそれぞれ等しい次元と容量を持っているものとする。

【0090】

また、機械学習を行う能力を持つノードは、図8に示す最下層(1層目)のエッジノードと最上位のクラウドノード(クラウドサーバと称してもよい)だけであるとし、クラウドノードは単体で機械学習を行い、エッジノードは全体で1つとなって機械学習を行うものとする。

【0091】

この場合、データのグルーピングについては、(1)クラウドノードに送られたデータ、及び、(2)エッジノードに送られたデータの2つのグループにグルーピングされる。また、各ノードは自身の所有するデータを比率 r_j でクラウドノードで処理させ、 $1-r_j$ でエッジノードに処理させる。これらの比率は $B_{jj}'G$ に対応する。

【0092】

10

20

30

40

50

また、ここでは、ネットワークの各階層ごとに転送レートは一律であるとする。よって、この例では、 r_1, r_2, r_3, r_c の4つのレートが存在する。

【0093】

図8のネットワーク構成において、各階層のリング1つあたりのノード数は、 $n_1=6, n_2=6, n_3=4$ である(n_1 が最下層の1層目、 n_2 が中間層の2層目、 n_3 が最上位層の3層目)。各層間エッジの転送速度を Q_{12} (1層目と2層目との間の転送速度)、 Q_{23} (2層目と3層目との間の転送速度)、 Q_{3c} (3層目とクラウドサーバとの間の転送速度)とし、各グループの機械学習速度を Re (エッジサーバグループ)、 Rc (クラウドサーバ)とする。また、1ノードから発生するMIBやSyslogのデータ量を1とする。

【0094】

この時、1層目のノードのクラウドノードへのデータ転送速度を t_{1u} とし、2層目のノードのクラウドノードへのデータ転送速度を t_{2u} とし、2層目のノードのエッジグループへのデータ転送速度を t_{2d} とし、3層目のノードのクラウドノードへのデータ転送速度を t_{3u} とし、3層目のノードのエッジグループへのデータ転送速度を t_{3d} とし、2層目のノードのエッジグループへのデータ転送速度を t_{cd} とする。

【0095】

この場合、実施例6において示した $_s$ の第一項は $\max\{t_{1u}, t_{2u}, t_{2d}, t_{3u}, t_{3d}, t_{cd}\}$ となる。一方で、エッジノードグループとクラウドノードで処理するデータ量をそれぞれ d_e, d_c とすると、 $_s$ の第2項は $\max_{\{\text{エッジorクラウド}\}}\{d_e/Re, d_c/Rc\}$ となる。

【0096】

また、モデルにはできるだけ多くの情報が含まれていた方が様々な次元間の関係性より異常を発見することができ、精度が上がると考えらるので、 $(\text{精度}) = (\text{エッジグループで処理するデータの次元数})^2 + (\text{クラウドで処理するデータの次元数})^2 / (\text{全データの総次元数})^2$ として精度を0.5~1の範囲で表すことにする。

【0097】

転送レート比率を $Q_{12}:Q_{23}:Q_{3c}=1:30:50$ とし、処理レート比率を $Re:Rc=10:15$ としてみると、実施例6の数式に従って転送時間と機械学習時間のそれぞれの最大値を取ったものが総処理時間となる。ここで $[r_1, r_2, r_3, r_c]=[1, 1, 1, 1]$ という全てのデータをクラウドサーバに送る場合を考える。この場合の結果は、図9に示すとおり、 $(\text{総処理時間})=54.76$ 、 $(\text{精度})=1$ となった。

【0098】

一方で、 $(\text{総処理時間}) \leq 35$ を満たすように、精度を最大化するようにレート $[r_1, r_2, r_3, r_c]$ を決定するという問題を解くと、図9に示すとおり、 $[r_1, r_2, r_3, r_c]=[0, 0.2, 0.1, 0.8]$ となり、 $(\text{総処理時間})=35, (\text{精度})=0.93$ となった。

【0099】

このように処理を分散させることで精度7%の低下に対し、36%処理速度を削減させることができ、処理速度と精度を双方鑑みて分散方法を決定するという本発明に係る技術の効果が確認された。

【0100】

(実施例のまとめ)

以上、説明したように、実施例1においては、ネットワーク外から得られるセンサデータと、ネットワーク内で得られるログデータに対してネットワーク上で機械学習による分析を行うために、ネットワーク上のノードに対して、データの処理を担う処理ノード、どの処理ノードがどのデータを処理するかについて制御を行う制御ノード、自身が得たデータを処理させるためにデータを処理ノードに転送する転送ノードの3種の機能を与える。

【0101】

実施例2においては、実施例1のネットワーク構成において、制御ノードが、各ノードの負荷状況に基づいて、各転送ノードがどの処理ノードにどのくらいの比率でどのデータを転送するかを制御する。また、機械学習において必要となるデータの正規化を行うために、制御ノードは、各処理ノード上に集約されたデータの統計量を集約計算し、正規化に

10

20

30

40

50

必要な情報を処理ノードへと再分配を行う。

【0102】

実施例3においては、実施例1のネットワーク構成において、制御ノードが、データ処理時間や機械学習の要求精度を目的関数もしくは制約条件として用いることで、転送ノードから収集されたデータをグルーピングし、各グループについて機械学習のモデルを生成する。また、グルーピングされた各モデルをどの処理ノードで訓練させるかについて、処理ノードのグルーピングも行う。また、訓練されたモデルを用いてテストを行う際に、どの処理ノードでテストさせるかについてもグルーピングを行う。この際テストを行う処理ノードグループは訓練の時に用いたものと違ってよい。

【0103】

実施例4においては、実施例1のネットワーク構成で機械学習としてネットワーク負荷情報の回帰を行う場合に、実施例2において、制御ノードに対して入力となる各ノードの負荷状況について、実際に計測した負荷ではなく、実施例3における機械学習によって回帰モデルで予測される負荷を入力とし、データ収集のサンプリングレートや、各データ転送ノードがデータ整形の前後でどの処理ノードにどのくらいの比率でデータを転送するかを制御する。

【0104】

実施例5は実施例2の詳細例である。実施例5では、制御ノードは各処理ノードの負荷情報を表す x_j (j : 処理ノードのラベル)及び各エッジの負荷情報を表す y_{ij} (i : 転送ノードのラベル)に依存する、各処理ノードのデータ整形速度 $P_j(x_j, \quad)$ (ここで \quad はデータの種類を表す)、各エッジのデータ伝送速度 $Q_{ij}(y_{ij})$ に基づきネットワーク全体での(生データ転送時間)+(データ整形時間)を最短にするように各転送ノードの各処理ノードへの整形前後のデータ転送レート A_{ij} (データ整形前転送レート)を決定する。

【0105】

実施例6は実施例3の詳細例である。実施例6では、訓練に用いるデータの期間を T_s 、テストに用いるデータの期間を T_p とした上で、機械学習の精度を G_d 、 T_s 、 T_p を変数とする関数 $F(G_d, T_s, T_p)$ として表し、指定された分析精度 \quad に対して $F(G_d, T_s, T_p) > \quad$ を満たすという条件を満たすように各処理ノードの機械学習の訓練速度 $R(x(g_{nk}), g_{dl})$ 及びテスト速度 $R'(x(g_{nk}), g_{dl})$ と $Q_{jj}'(y_{jj}')$ に基づき(整形済みデータのデータ転送時間)+(訓練もしくはテストにかかる時間)を最小にするように訓練に用いるデータの期間 T_s 、テストに用いるデータの期間 T_p 、データ整形後転送レート $B_{jj}'G$ (j' は機械学習を行う処理ノード)、データのグルーピング G_d 、及び処理ノードのグルーピング G_n の最適化を行う。

【0106】

実施例7も実施例3の詳細例である。実施例7では、訓練及びテストを行う時間間隔をそれぞれ \quad_s 、 \quad_p とし、訓練及びテストに要する時間が \quad_s 及び \quad_p より小さくなるような制約条件の元で、実施例6における精度 $F(G_d, T_s, T_p)$ を最大化する。

【0107】

実施例8は、実施例4の詳細例である。実施例8では、周期 \quad_s 、 \quad_p で各ノードの状態 x_j 及び各エッジの状態 y_{ij} 、 y_{jj}' に対する回帰の訓練とテストを行っている場合において、ある周期の訓練やテストが始まる前の1周期前に行われたテストの結果で得られた回帰予測値 $x_{p_i}(t - \quad_p)$ (各処理ノードの状態)、 $y_{p_{ij}}(t - \quad_p)$ 及び $y_{p_{jj}'}(t - \quad_p)$ (各エッジの状態)に基づいて A_{ij} 、 G 、 $B_{jj}'G$ 、 T_s 、 T_p を制御する。

【0108】

実施例9は、実施例5、6、7のそれぞれに適用可能な例である。実施例5、6、7において、転送ノードから処理ノードへのデータ転送量をできるだけ抑えるために、制御ノードは、レート A_{ij} 、 B_{jj}' に関する制約条件 $H(A_{ij}, B_{jj}')(\quad)$ の元で A_{ij} 、 $B_{jj}'G$ を決定する。

【0109】

上述したとおり、本発明の実施の形態により、データ転送機能を有するノードである転

10

20

30

40

50

送ノードと、データ処理機能を有するノードである処理ノードと、制御機能を有するノードである制御ノードを有し、機械学習による処理を実行するネットワークにおける前記制御ノードとして機能する制御装置であって、各処理ノードの負荷情報に基づいて、各転送ノードの各処理ノードへのデータの転送レートを決定する第1決定手段と、データのグルーピングと処理ノードのグルーピングに基づいて、データを処理する処理ノードを決定する第2決定手段とを備えることを特徴とする制御装置が提供される。

【0110】

転送比率計算部106は、第1決定手段及び第2決定手段の例である。

【0111】

前記第1決定手段は、各処理ノードの負荷情報と、転送ノードと処理ノード間の負荷情報とに基づいて、機械学習のためのデータ整形前の生データ転送時間と、処理ノードで実行されるデータ整形の時間との和が最小になるように、前記転送レートを決定することとしてもよい。

10

【0112】

前記第1決定手段は、前記負荷情報として、機械学習による回帰モデルで予測される負荷情報を使用することとしてもよい。

【0113】

前記第2決定手段は、処理ノード間の負荷情報と、各処理ノードでの機械学習の処理速度とに基づいて、機械学習の精度が所定の条件を満たすように、機械学習のための整形済みデータの転送時間と機械学習の処理にかかる時間との和が最小になるように、データを

20

【0114】

前記第2決定手段は、処理ノード間の負荷情報と、各処理ノードでの機械学習の処理速度とに基づいて、機械学習のための整形済みデータの転送時間と機械学習の処理にかかる時間との和が所定の条件を満たすように、機械学習の精度が最大になるように、データを処理する処理ノードを決定することとしてもよい。

【0115】

前記第1決定手段は、データ転送量に関する制約条件を満足するように前記転送レートを決定することとしてもよい。

【0116】

また、本実施の形態により、コンピュータを、上記制御装置における各手段として機能させるためのプログラムが提供される。

30

【0117】

また、本実施の形態により、データ転送機能を有するノードである転送ノードと、データ処理機能を有するノードである処理ノードと、制御機能を有するノードである制御ノードを有し、機械学習による処理を実行するネットワークにおける前記制御ノードとして機能する制御装置が実行するデータ処理制御方法であって各処理ノードの負荷情報に基づいて、各転送ノードの各処理ノードへのデータの転送レートを決定する第1決定ステップと、データのグルーピングと処理ノードのグルーピングに基づいて、データを処理する処理ノードを決定する第2決定ステップとを備えることを特徴とするデータ処理制御方法が提

40

【0118】

以上、本実施の形態について説明したが、本発明はかかる特定の実施形態に限定されるものではなく、特許請求の範囲に記載された本発明の要旨の範囲内において、種々の変形・変更が可能である。

【符号の説明】

【0119】

101 入力部

102 記憶部

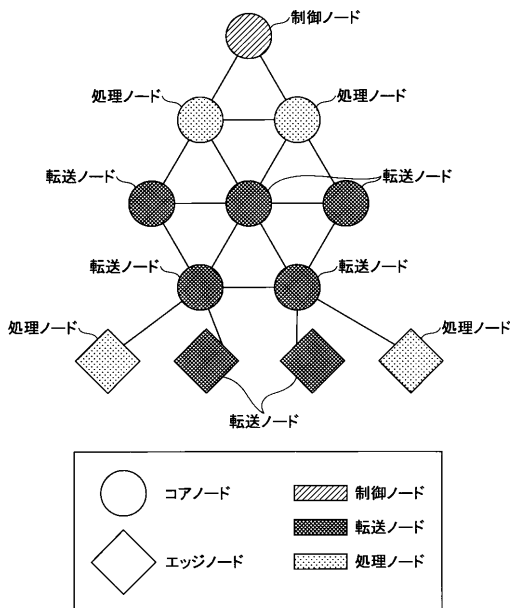
103 モデル計算部

50

- 1 0 4 正規化パラメータ管理部
- 1 0 5 機械学習テスト部
- 1 0 6 転送比率計算部
- 1 0 7 機械学習訓練部
- 1 0 8 出力部
- 1 5 0 ドライブ装置
- 1 5 1 記録媒体
- 1 5 2 補助記憶装置
- 1 5 3 メモリ装置
- 1 5 4 C P U
- 1 5 5 インターフェース装置
- 1 5 6 表示装置
- 1 5 7 入力装置

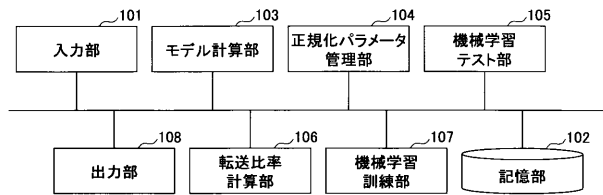
【 図 1 】

本発明の実施の形態におけるシステム構成の例を示す図



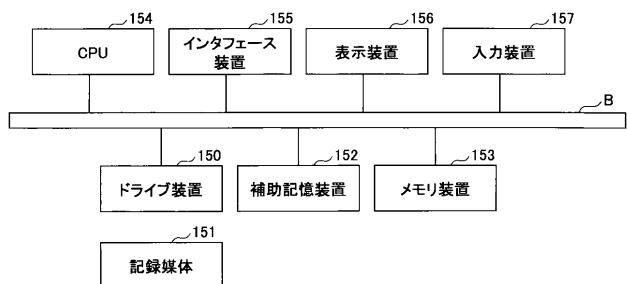
【 図 2 】

制御装置の機能構成の一例を示す図



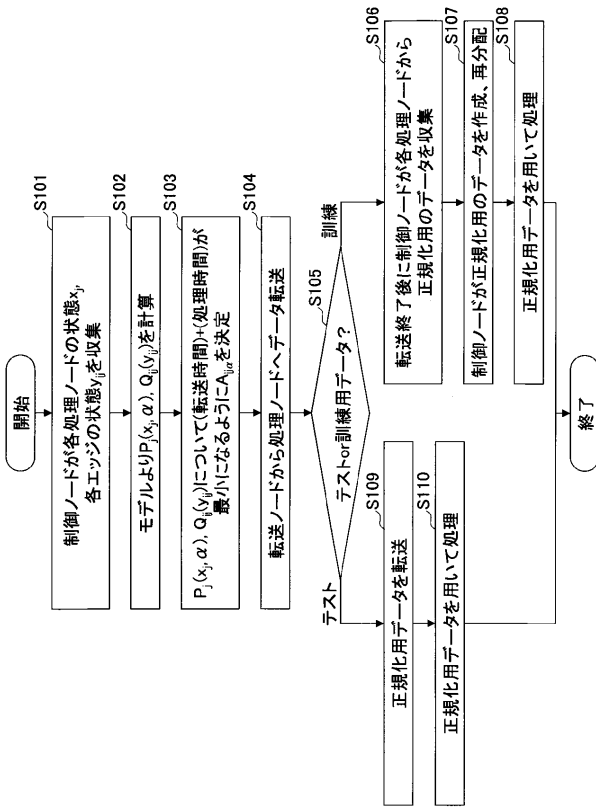
【 図 3 】

制御装置のハードウェア構成の一例を示す図



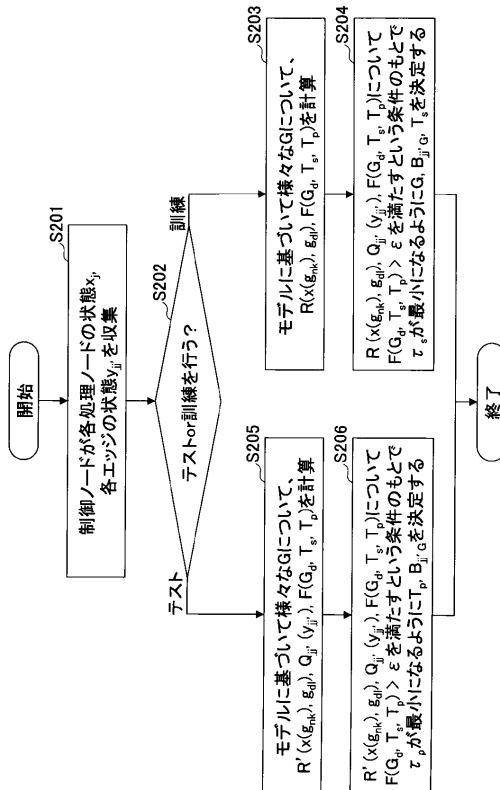
【図4】

実施例5における動作例を説明するためのフローチャート



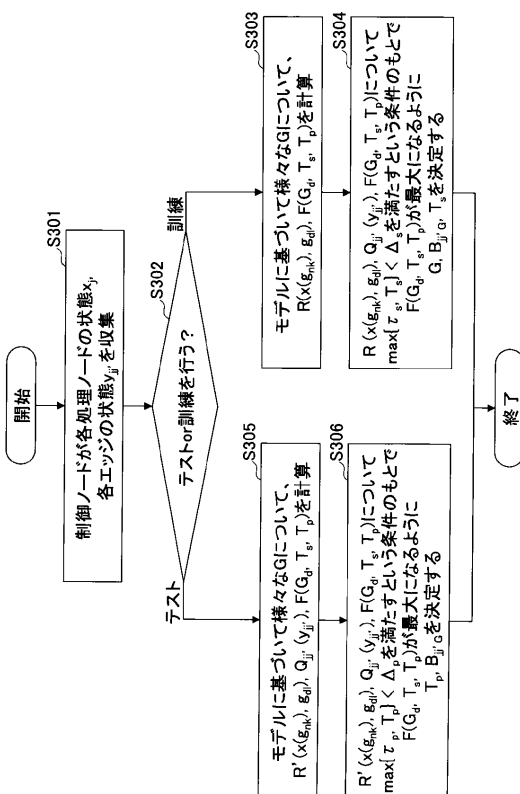
【図5】

実施例6における動作例を説明するためのフローチャート



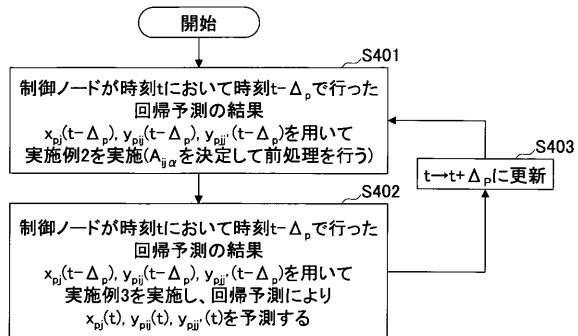
【図6】

実施例7における動作例を説明するためのフローチャート



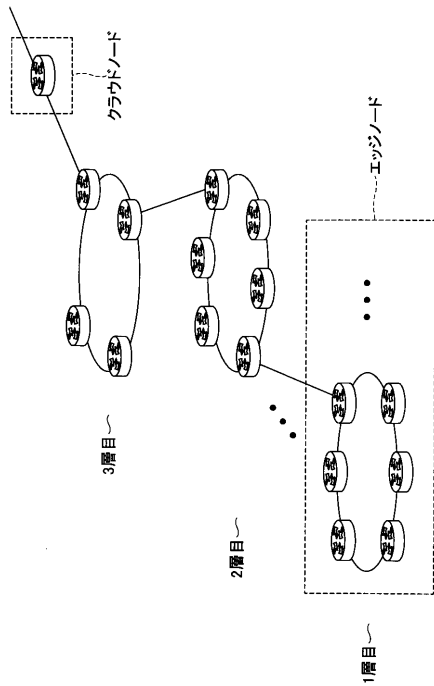
【図7】

実施例8における動作例を説明するためのフローチャート



【 図 8 】

効果を説明するためのラダー型ネットワークの構成例を示す図



【 図 9 】

クラウドエッジ対応による処理速度と精度の関係性を示す図

	クラウドサーバ のみでの処理	速度を制限した 精度最適化(速度<35)	削減比率
レート [r1,r2,r3,rc]	[1, 1, 1, 1] (全データクラウド処理)	[0, 0.2, 0.1, 0.8] (1,2,3層目⇒エッジ処理) (クラウド⇒クラウド処理)	
処理時間	54.76	35	36%
精度	1	0.93	7%

フロントページの続き

- (72)発明者 池田 泰弘
東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
- (72)発明者 川原 亮一
東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
- (72)発明者 新熊 亮一
京都府京都市左京区吉田本町3番地1 国立大学法人京都大学内