

【公報種別】特許法第 17 条の 2 の規定による補正の掲載
 【部門区分】第 6 部門第 3 区分
 【発行日】平成26年9月4日 (2014.9.4)

【公表番号】特表2014-510342(P2014-510342A)
 【公表日】平成26年4月24日 (2014.4.24)
 【年通号数】公開・登録公報2014-021
 【出願番号】特願2013-555715(P2013-555715)
 【国際特許分類】

G 0 6 N 3/00 (2006.01)

G 0 6 F 9/45 (2006.01)

【F I】

G 0 6 N 3/00 5 5 0 A

G 0 6 F 9/44 3 2 0 A

【手続補正書】
 【提出日】平成26年7月18日 (2014.7.18)
 【手続補正 1】
 【補正対象書類名】特許請求の範囲
 【補正対象項目名】全文
 【補正方法】変更
 【補正の内容】
 【特許請求の範囲】
 【請求項 1】

MapReduce 環境で機械学習 (ML) アルゴリズムを処理するためのコンピュータによって実行される方法であって、

前記MapReduce 環境で実行されるMLアルゴリズムを受け取るステップと、

前記MLアルゴリズムをシーケンス内の複数のステートメント・ブロックに構文解析するステップであって、それぞれのステートメント・ブロックが複数の基本演算 (hop) を含む、ステップと、

それぞれのステートメント・ブロックの実行計画を自動的に決定するステップであって、前記実行計画のうちの少なくとも 1 つが 1 以上の下位演算 (lop) を含み、それぞれのステートメント・ブロックの実行計画を決定するステップが、hop を少なくとも 1 つの lop に変換するステップを含む、ステップと、

前記複数のステートメント・ブロックの前記シーケンスで前記実行計画を実施するステップと

を含む方法。

【請求項 2】

前記ステートメント・ブロックのうちの少なくとも 1 つの前記実行計画を決定するステップが、

前記ステートメント・ブロックを解析して、前記複数の hop 間の関係 (interconnections) を決定するステップと、

前記関係を使用して、行列およびスカラーのうちの少なくとも 1 つにおける前記複数の hop の無閉路有向グラフ (HOPDag) を作成するステップと
 を含む、請求項 1 に記載のコンピュータによって実行される方法。

【請求項 3】

それぞれのステートメント・ブロックの前記実行計画を決定するステップが、前記 HOPDag を複数の lop を含む少なくとも 1 つの下位物理実行計画 (LOPDag) に変換するステップをさらに含む、請求項 2 に記載のコンピュータによって実行される方法。

【請求項 4】

前記 L O P D a g が、スカラー演算の表現をさらに含む、請求項 3 に記載のコンピュータによって実行される方法。

【請求項 5】

前記 h o p を少なくとも 1 つの l o p に変換するステップが、

前記 h o p と同等の複数の l o p のグループを決定するステップと、

コストベースの最適化を使用して、前記複数の l o p のグループから、前記 l o p のグループのうちの少なくとも 1 つの関連コストよりも低い関連コストを有する 1 つの l o p のグループを選択するステップと

を含む、請求項 1 に記載のコンピュータによって実行される方法。

【請求項 6】

それぞれのステートメント・ブロックの前記実行計画を決定するステップが、前記 L O P D a g を M a p R e d u c e ジョブの代表的なワークフローに変換するステップをさらに含み、前記 L O P D a g を変換するステップが、複数の l o p を 1 つの M a p R e d u c e ジョブによって表すことができるかどうかを判定するステップを含む、請求項 3 に記載のコンピュータによって実行される方法。

【請求項 7】

それぞれのステートメント・ブロックの前記実行計画を決定するステップが、前記 H O P D a g を最適化するステップをさらに含み、前記 H O P D a g の最適化が、代数的な書換えのうちの少なくとも 1 つおよび複数の中間行列の物理表現の選択を含む、請求項 6 に記載のコンピュータによって実行される方法。

【請求項 8】

それぞれのステートメント・ブロックの前記実行計画を決定するステップが、コストベースの最適化を使用することによって、H O P D a g を L O P D a g に変換している間に前記 L O P D a g を最適化するステップをさらに含む、請求項 7 に記載のコンピュータによって実行される方法。

【請求項 9】

前記実行計画を実施するステップが、制御プログラムおよびクラスタ上の M a p R e d u c e ジョブの複数のワークフローによってスカラー演算の実行を管理するステップを含む、請求項 1 に記載のコンピュータによって実行される方法。

【請求項 10】

M a p R e d u c e ジョブの前記複数のワークフローの実行が、前記ワークフローが作成されたステートメント・ブロックのタイプを使用して、前記ワークフローを実行の際にスキップすべきか、または実行の際に繰り返すべきかを判定するステップを含む、請求項 9 に記載のコンピュータによって実行される方法。

【請求項 11】

複数の l o p を 1 つの M a p R e d u c e ジョブによって表すことができるかどうかを判定するステップをさらに含み、そのような判定が、前記 l o p の特性に基づいて複数の l o p を 1 つの M a p R e d u c e ジョブにピギーバックするステップを含む、請求項 1 に記載のコンピュータによって実行される方法。

【請求項 12】

前記 M a p R e d u c e ジョブを作成するステップをさらに含み、前記 M a p R e d u c e ジョブが、それぞれのフェーズがピギーバック方法に従って複数の l o p を実行するようにパラメータ化された、少なくとも 3 つのフェーズに分割される汎用 M a p R e d u c e ジョブからインスタンス化される、請求項 11 に記載のコンピュータによって実行される方法。

【請求項 13】

前記 M L アルゴリズムを実施するためのデータ・ブロック表現のサイズを最適化するステップを含む前記実行計画を最適化するステップをさらに含み、請求項 1 に記載のコンピュータによって実行される方法。

【請求項 14】

前記データ・ブロック表現のサイズを最適化する前記ステップが、それぞれのデータ・ブロックの局所疎性に依存する、請求項13に記載のコンピュータによって実行される方法。

【請求項15】

コンピュータ可読プログラムを記憶するためのコンピュータ使用可能記憶媒体を含むコンピュータ・プログラム製品であって、前記コンピュータ可読プログラムが、コンピュータ上で実行されるとき、前記コンピュータに、MapReduce環境で機械学習（ML）アルゴリズムを処理するための操作であって、

前記MapReduce環境で実行されるMLアルゴリズムを受け取るステップと、

前記MLアルゴリズムをシーケンス内の複数のステートメント・ブロックに構文解析するステップであって、それぞれのステートメント・ブロックが複数の基本演算（hop）を含む、ステップと、

それぞれのステートメント・ブロックの実行計画を自動的に決定するステップであって、前記実行計画のうちの少なくとも1つが1以上の下位演算（lop）を含み、それぞれのステートメント・ブロックの実行計画を決定するステップが、hopを少なくとも1つのlopに変換するステップを含む、ステップと、

前記複数のステートメント・ブロックの前記シーケンスで前記実行計画を実施するステップと、

複数のlopを1つのMapReduceジョブによって表すことができるかどうかを判定するステップであって、そのような判定が、前記lopの特性に基づいて複数のlopを1つのMapReduceジョブにビジーバックする、前記判定するステップと、を含む前記操作を実行させる、コンピュータ・プログラム製品。

【請求項16】

前記コンピュータ可読プログラムが、前記コンピュータに、前記ステートメント・ブロックのうちの少なくとも1つの前記実行計画を決定するための操作であって、

前記ステートメント・ブロックを解析して、前記複数のhop間の関係を決定するステップと、

前記関係を使用して、行列およびスカラーのうちの少なくとも1つにおける前記複数のhopの無閉路有向グラフ（HOPDag）を作成するステップとを含む前記操作を実行させる、請求項15に記載のコンピュータ・プログラム製品。

【請求項17】

前記コンピュータ可読プログラムが、前記コンピュータに、それぞれのステートメント・ブロックの前記実行計画を決定するための操作であって、前記HOPDagを複数のlopを含む少なくとも1つの下位物理実行計画（LOPDag）に変換するステップを含む前記操作を実行させる、請求項16に記載のコンピュータ・プログラム製品。

【請求項18】

前記LOPDagが、スカラー演算の表現をさらに含む、請求項15に記載のコンピュータ・プログラム製品。

【請求項19】

前記コンピュータ可読プログラムが、前記コンピュータに、前記HOPDagを少なくとも1つのLOPDagに変換するための操作であって、前記HOPDagで実施されるそれぞれのhopについて、前記hopを少なくとも1つのlopに変換するステップを含む前記操作を実行させる、請求項18に記載のコンピュータ・プログラム製品。

【請求項20】

前記コンピュータ可読プログラムが、前記コンピュータに、それぞれのステートメント・ブロックの前記実行計画を決定するための操作であって、前記lopをMapReduceジョブに変換するステップを含む前記操作を実行させ、前記lopを変換するステップが、複数のlopを1つのMapReduceジョブによって表すことができるかどうかを判定するステップを含む、請求項19に記載のコンピュータ・プログラム製品。

【請求項21】

前記コンピュータ可読プログラムが、前記コンピュータに、前記複数の `loop` を1つの `MapReduce` ジョブによって表すことができるかどうかを判定するための操作であって、`loop` を1つの `MapReduce` ジョブにピギーバックするステップを含む前記操作を実行させる、請求項 20 に記載のコンピュータ・プログラム製品。

【請求項 22】

前記コンピュータ可読プログラムが、前記コンピュータに、それぞれのステートメント・ブロックの前記実行計画を決定するための操作であって、前記 `HOP DAG` を最適化するステップを含む前記操作を実行させ、前記 `HOP DAG` の最適化が、代数的な書換えのうちの少なくとも1つおよび複数の中間行列の物理表現の選択を含む、請求項 19 に記載のコンピュータ・プログラム製品。

【請求項 23】

前記コンピュータ可読プログラムが、前記コンピュータに、それぞれのステートメント・ブロックの前記実行計画を決定するための操作であって、`HOP DAG` を `LOPDAG` に変換している間に前記 `LOPDAG` を最適化するステップを含む前記操作を実行させ、そのような変換が、

前記 `hop` と同等の複数の `loop` のグループを決定するステップと、

コストベースの最適化を使用して、前記複数の `loop` のグループから、前記 `loop` のグループのうちの少なくとも1つの関連コストよりも低い関連コストを有する1つの `loop` のグループを選択するステップと

を含む、請求項 19 に記載のコンピュータ・プログラム製品。

【請求項 24】

`MapReduce` 環境で機械学習 (ML) アルゴリズムを処理するためのシステムであって、

前記 `MapReduce` 環境で実行される ML アルゴリズムを受け取るための手段と、

前記 ML アルゴリズムをシーケンス内の複数のステートメント・ブロックに構文解析するための手段であって、それぞれのステートメント・ブロックが複数の基本演算 (`hop`) を含む、手段と、

それぞれのステートメント・ブロックの実行計画を自動的に決定するための手段であって、前記実行計画のうちの少なくとも1つが1以上の下位演算 (`loop`) を含む、手段と、

前記複数のステートメント・ブロックの前記シーケンスで前記実行計画を実施するための手段と、

前記 ML アルゴリズムを実施するためのデータ・ブロック表現のサイズを最適化する手段を含む前記実行計画を最適化する手段と、

を含むシステム。

【請求項 25】

それぞれのステートメント・ブロックの前記実行計画を決定するための手段が、行列、ベクトル、およびスカラーのうちの少なくとも1つにおける前記複数の `hop` の無閉路有向グラフ (`HOP DAG`) を作成するための手段を含み、

それぞれのステートメント・ブロックの前記実行計画を決定するための手段が、前記 `HOP DAG` を複数の `loop` を含む少なくとも1つの下位物理実行計画 (`LOPDAG`) に変換するための手段を含み、

前記 `HOP DAG` を少なくとも1つの `LOPDAG` に変換するための手段が、前記 `HOP DAG` で実施されるそれぞれの `hop` について、前記 `hop` を少なくとも1つの `loop` に変換するための手段を含み、

それぞれのステートメント・ブロックの前記実行計画を決定するための手段が、前記 `loop` を `MapReduce` ジョブの複数のワークフローに変換するための手段を含み、前記 `loop` を変換する手段が、複数の `loop` を1つの `MapReduce` ジョブにピギーバックする手段を含む、

請求項 24 に記載のシステム。