

計算パラメータ推薦システムのフレームワークについて

白 山 晋[†] 齋 藤 幸二 郎^{††}
竹 森 恵 一^{††} 太 田 高 志^{†††}

ユーザが目的に適した計算プログラムを選択し、問題解決のために最適な計算パラメータを決定するための支援法を考察し、計算パラメータだけではなく、物理パラメータ、設計パラメータなどのパラメータスタディに対しても有効な最良パラメータの探索法とルール抽出法、そして、推薦システムのフレームワークを提案した。その実効性を確かめるために、2つの試作システムを作り、数値実験を行った。その結果、パラメータ推薦システムの基本概念の有用性が確認できた。また、システムにおける計算部の並列性は高く、通信量が少ないために粗結合の並列環境であっても効率的なパラメータ推薦が可能であること、データの蓄積・管理を含めて広域分散計算に適したアプリケーションであることを示せた。

A Framework of Recommender System for Choices of Computational Parameters

SUSUMU SHIRAYAMA,[†] KOJIRO SAITO,^{††} KEIICHI TAKEMORI^{††}
and TAKASHI OHTA^{†††}

Recently, we can access many kinds of CAE codes. However, selecting a suitable code for the purpose of the effective and efficient computations is not easy for users have not knowledge of CAE and no concern about the feature of computer system. In additions, it is quite difficult to determine the appropriate computational parameters even if a well-established CAE code is utilized. In this study, we propose a framework for the selection process of the parameters using distributed resources such as computer, database and knowledge. Two prototype systems are implemented on a PC cluster and a Grid. Numerical experiments show that our proposed framework will be valid to recommend an appropriate computational parameter for users, and to extract unknown rules in the computations.

1. 緒 言

CAEの発展とともに、多くの計算プログラムが生み出されている。技術計算プログラムは、制御構造が単純であるという特徴を持ち、構造化されているものが多い。解析手順の大枠が確立されているためである。つまり、独立、従属変数の配列化、支配方程式の離散化、時間発展、あるいは反復などによる配列変数の更新を適切な初期値、境界値の付与によって行い、決められたタイミングで変数の値を解析するという手順である。また、多くのプログラムは線形ソルバを核として構成することができる^{1)~3)}。したがって、独立、従

属変数のデータ構造が明確であれば、ある部分の機械的変更によって最新の基礎的研究の成果を取り入れることも難しくはない。また、並列計算などの計算環境の多様化に対しても定石ともいえる対応策がある。

プログラム群を統合化する際にもこの性質が重要な役割を果たしている。たとえば、線形ソルバ部のライブラリの扱いによる計算環境の隠蔽、メタレベルでの計算環境の隠蔽⁴⁾等によって、プログラム群から解析システムへと発展している。しかしながら、個々の問題に着目した場合、従来の個別プログラムの書き換えという方法以上の効果が現れていないという指摘もある。構造化計算との結合といっても、配列変数の共通化によって作業の簡素化は可能であるし、オブジェクト指向などの新しいプログラミングスタイルを利用しなくても、機能的には同等のプログラムを書くことができる。システム化の一方で個別のプログラムが淘汰されることなく、増え続けている理由である。また、強い非線形現象を扱う分野においては、問題ごとに計

[†] 東京大学人工物工学研究センター

RACE, The University of Tokyo

^{††} 東京大学大学院工学系研究科

School of Engineering, The University of Tokyo

^{†††} IBM 東京基礎研究所

Tokyo Research Laboratory, IBM

算アルゴリズムレベルでの修正が必要である。それらはモジュールの選択という形で個別プログラムや解析システムに吸収されている場合が多い。その結果、

- ユーザが目的に適した計算プログラムを選択することは難しい、
- あるプログラムが選択できたとしても、問題解決のために最適な計算パラメータを決定することが容易ではない、

という問題が生じ、ユーザの負担を軽減するための仕組みが模索されている。いくつかのソフトウェアは、知識ベースを核としたパラメータ選択の支援法を採用している^{5),6)}。しかしながら、効率的、かつ高精度な計算を行うためには、計算対象に対する知見とともに、数値計算に対する高度な専門知識が必要になる。機械的に処理できる条件は少なく、実用的な知識ベースにはなっていない。

本稿では、ユーザが目的に適した計算プログラムを選択し、問題解決のために最適な計算パラメータを決定するための支援法を考察し、計算パラメータだけではなく、物理パラメータ、設計パラメータなどのパラメータスタディに対しても有効な最良パラメータの探索法とルール抽出法、そして、推薦システムを提案する。また、計算パラメータの探索空間は大きく、組合せ的爆発が予想されるため、探索を効率化する方法についても検討する。

2章でパラメータ推薦システムの枠組みを検討し、3章で必要となる構成要素を提案する。4章で利用する計算環境の考察を行う。5章で2つの試作システムによる数値実験の結果を示す。6章はまとめである。

2. パラメータ推薦システム

パラメータスタディの自動実行による最良パラメータの探索が提案する推薦システムの核である。これに、

- 高効率、高精度の推薦
- 知識の自動抽出

という2つの機能を加える。知識抽出の主たる目的は最良パラメータ探索のための知識形成ではあるが、評価値群からのものも含めることにより拡張性を高める。

2.1 パラメータスタディ

CAEにおけるパラメータは主として、

- 設計パラメータ
- 物理パラメータ
- 計算パラメータ

である。一般的にパラメータスタディという場合は、設計パラメータや物理パラメータを系統的に変えて計算し、計算結果群を算出することを意味する。目的は、

計算結果群の取得そのものである場合と、評価値を決めて何かしらのパラメータを探索する場合に大別される。提案するシステムの目指すものは、問題が与えられたときに適切な計算プログラムと計算パラメータを推薦するという点で、後者のパラメータスタディと同じである。ただし、計算結果群とパラメータの関係も利用するので前者の意味合いも含む。

パラメータスタディにおいて最も重要なことの1つが効率化である。並列計算、あるいは分散計算が一般的な手法ではあるが、既存の計算結果の利用、既存の知見や解析者の持つ経験と勘による戦略的なパラメータの絞り込みも効率化に寄与する。既存の計算結果が蓄えられているのであれば、パラメータを検索キーとして、それを探し、利用すればよい。しかしながら、計算結果には再現性の問題がある。同一プログラム、同一パラメータであっても、計算環境が異なれば計算データに多少の違いが生じるように、一意に対応づけられたと考えられるパラメータとデータにも属性情報が不足している場合も少なくない。また、仮にパラメータと一意に対応がつくデータであっても、必要とする情報を含んでいるとは限らない。逆にパラメータに多少の違いがあっても、必要とされる情報を含む既存の結果が存在する可能性もある。

既存の知見や解析者の持つ経験と勘によるパラメータの絞り込みは一過性で行われることが多く、そのままでは個人の暗黙知が増えるだけである。普遍性を追求するためには、暗黙知の形式知化、機械処理可能な知識の形成が必要である。これは結果の判断の際にもいえることである。人的資源を活用するための具体的な方法論が不足している。

パラメータスタディの目的が最適化問題の解を求める場合、いかに少ないパラメータで最適値を求めるかが重要であり、最終的な結果が成果であるため、履歴情報は棄却されることが多い。結果となるデータ群が分析されることはあるが、履歴情報の中から知識を抽出する試みは少ない。

以上をまとめると、パラメータスタディの課題は、

- (A) 既存の計算結果(データ)を直接利用するための検索技術
- (B) 個々の熟練者によって蓄えられている知見の活用、あるいは熟練者自身(人的資源)の利用法
 - (B1) 暗黙知の形式知化
 - (例) 数理的に分かっていないルールの抽出
 - (B2) 機械処理できる形での知識形成
- (C) 履歴情報の利用
- (D) (A)~(C)を実現するための方法論、システム化

が不足していることと考えられる。

2.2 パラメータマイニング

パラメータスタディにおける課題の中で知識の形成に対して有効な方法の1つは、データマイニングの利用である。

一般的な意味でのデータマイニングは、データの中からの知識抽出を意味する。ただし、一方向的な探索ではなく、得られた知識、あるいはルールから探索データを選別するといったフィードバックプロセスを含む(図1上)。このプロセスもパラメータスタディの効率化に利用できる。データマイニングの問題は、データそのものが探索対象であり、非常に大きい探索空間の中で知識抽出が行われることである。この問題を解決する方法の1つはデータの構造化である。データを何かしらのクラスに仕分けることができれば、階層的な探索が可能となり探索効率が向上する。また、探索対象となるデータの所属属性を利用するドメインモデルの導入も効果的である。具体的には、データ生成プロセスからの情報を利用することになる。

一方、パラメータを介したデータ群を考えると、データはパラメータによってクラスに分けられ、データ生成プロセスとも関連づけられている。そこで、データと知識抽出の間にパラメータ空間を置き、パラメータ空間を介してデータマイニングを行う(図1下)。これをパラメータマイニングと呼ぶことにする。パラメータマイニングでは、目的に適したパラメータを探索しながらルール抽出を試みる。パラメータとデータの関係から探索空間は階層的な構造を持ち、データ生成プロセスからの情報を利用したドメインモデルの導入も容易である。ただし、先述したようにパラメータがデータの内容を反映していない場合には工夫が必要になる。

2.3 推薦システムの枠組み

提案する推薦システムは、パラメータスタディとパラメータマイニングを組み合わせたものである。具体的なシステムを考えるうえで、2.1節の課題:(A)~(D)に対する答えとデータ生成プロセスからの情報の利用が重要である。繰返しになるが、

- 探索空間の更なる階層化
- 人的資源の利用と機械処理
- ドメインモデルの利用

が鍵となる。属性情報の不足によりパラメータとデータが一意に対応づけられていない場合やパラメータと複数の評価値が対応している場合には、データからのマイニングが必要となり、効率的、かつ高精度のパラメータ推薦が難しくなる。属性情報のパラメータ化、

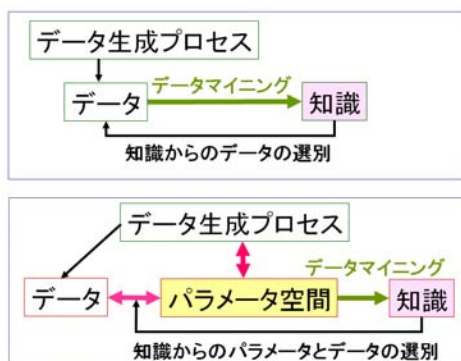


図1 データマイニング(上)とパラメータマイニング(下)の違い
Fig. 1 Difference between data mining (upper figure) and parameter mining (lower figure).

欠損情報の補完、パラメータの階層化が必要になる。また、システム全体で機械処理に適した情報の記述を考える必要がある。さらに着目しているドメインからの情報の利用を考えると、データ生成プロセスとの連携強化が重要であり、知識抽出に適したデータ生成を行うための仕組みが必要になる。これらは構成要素の中で考察すべき課題である。

次にパラメータ推薦の手順を考える。提案システムでは、

- 問題設定と計算に対する要求項目の決定
- パラメータ空間の生成
- 評価関数の設定
- 最適化問題としての定式化
- ルールの抽出
- 既存知識の利用と新たな知見の保存と再利用
- 履歴の分析

という一連の手続きによってパラメータ推薦が行われる。

はじめに、問題設定と要求項目に従って、パラメータ空間と評価関数を定義する。評価関数は、ある部分の誤差を最小にする、発散しない、計算時間が短いなど、目的に応じて設定する。評価関数に関しては、パラメータと評価値の関係を陽に表す必要はなく、パラメータが決まれば評価値が計算できるという仕組みがあればよい。評価関数が決まると、ある種の最適化問題を設定し、それを解く。その過程の中でルール抽出を行う。一連のプロセスに探索空間を制限するための既存知識の利用や、新たな知見の再利用などのフィードバックプロセスを加える。

以上がパラメータ推薦システムの大枠である。

表 1 CFD プログラムのパラメータ化

Table 1 Parameterization of the CFD programs.

code	用途				離散化		対流項
	圧縮	熱	2D	3D	FDM	FEM	2nd
a-flow	0	1	1	1	1	0	1
b-flow	1	1	1	1	0	1	0

3. パラメータ推薦システムの構成要素

推薦システムの構成要素全体で重要なことは、できる限り多くの作業を機械処理で行えるようにすることである。たとえば、機械処理が難しい対象を符号化によってパラメータ化する、セマンティックウェブ¹⁶⁾のように機械処理を可能とする情報の構造を与えるなどの方法が考えられる。

3.1 パラメータ空間

同一の計算対象に対して適用可能な計算プログラムは多く存在する。はじめに計算プログラムをいくつかの項目に従って符号化し、第 1 のパラメータ空間を形成する。たとえば、プログラムに備わる機能の有無を、0 と 1 で表すことによって、計算プログラムをパラメータによって表現する。

表 1 に示すようにプログラム b-flow は、1111010 と表される。このようにすれば、データがプログラムのどの機能を用いて生成されたのかも分かる。つまり、生成されたデータとデータ生成プロセスの連携を強化するためのデータ生成プロセスのパラメータ化になりうる。次に、あるプログラムが選択された場合の計算パラメータを第 2 のパラメータ空間として定義する。さらに計算環境を第 3 のパラメータ空間とし、すべてのパラメータ空間の中から最良パラメータを探索する。ただし、一般的には、計算結果とパラメータは完全に対応していない。たとえば、プログラムの符号化は標準化されたものではない。組織ごとに異なる項目が存在する可能性もあれば、同じ項目であっても違うものを示していることもある。セマンティックウェブのオントロジー層による吸収が 1 つの解決策である。また、パラメータの一部分によって計算結果の典型的なパターンを示している場合も想定できる。この場合は、計算結果の内容検索によって対処する。このようなパラメータ空間と計算結果を完全に対応させるための仕組みも推薦システムの構成要素として必要になる。

3.2 最適化問題への置き換えと探索過程

パラメータ空間を拡大すると系統的な絞り込みが必要になる。問題解決に適したパラメータを効率的に見つけるためには、数値計算理論などの方法によって制約条件を積み重ね、探索空間を縮小することが望まし

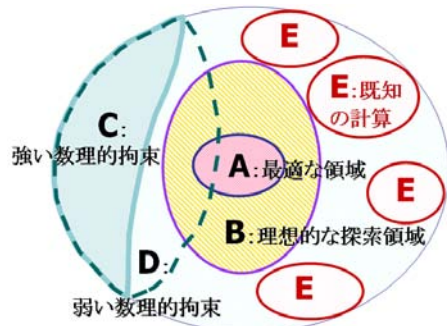


図 2 探索空間の構造

Fig. 2 Structure of a search space.

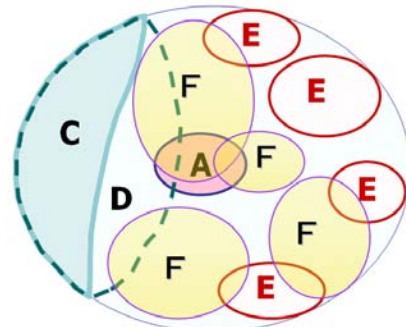


図 3 実際の最適値探索の模式図

Fig. 3 Illustration of an actual searching for optimal value.

い。たとえば、CFD であれば、乱流モデルの研究などの流体理論からのパラメータの限定や、計算スキームのさらなる高精度化・高速化によるスキームの淘汰、計算の安定性条件のような数理的考察が必要になる。これらの方法によって、パラメータがとりうる範囲が限定され、探索空間を縮小できる。理想的な探索を模式的に示したものが図 2 である。

図 2 は、パラメータ空間全体を示している。図中 C は、CFL 条件のような数理的制約条件において、機械処理によって探索を除外できる領域を示している。また、この領域のことを強い数理的拘束領域と呼ぶことにする。D は、数理的制約条件は存在しているが、問題固有であったり、明示的でなかったり、機械処理ができなかったりする領域を示す。最良パラメータの探索を、ある評価値に対する最適化問題と見なすと、既知の計算領域 E を除外しながら、最適な領域 A の近傍 B から探索することが望ましい。パラメータ空間の性質によってはこのような探索によって最適化を行う方法は存在する。しかしながら、実際は、図 3 の F のように最良パラメータを含む領域近傍ではあるが既存の計算結果の検索に失敗するものや、局所解に捕

られるものなど、理想的な最適化が行われない場合の方が多だろう。

提案するシステムでは、最良パラメータの探索過程を知識抽出に利用する。具体的には、知識抽出を

- (a) 拘束条件の明示（機械処理可能な表現）
- (b) 拘束条件を含む領域の識別
- (c) 探索履歴からのルール抽出

のように細分化して考える。(a)は、探索過程の中で弱い数理的拘束領域に属するパラメータと評価値から何かしらのルール抽出を行うことを示す。(b)で考えられるのは機械学習である。ルールが見つけれなくても、たとえば、発散解を与えるパラメータと収束解を与えるものを用いて、教師付き学習を行えば、あるパラメータを与えたときの予測値からどの領域のパラメータが判断する情報が得られる。(c)は、探索の履歴情報を分析することで、何かしらの戦略をルールとして抽出する試みである。

3.3 遺伝的アルゴリズムによる最適化

2.3節で示したように提案する枠組みにおいては、形式的に最適化問題を解くことになる。ただし、前節で述べたように最適値を求めることよりも、探索過程で得られる情報を利用して機械学習やルール抽出を行うことが重要になる。このため、大域的、独立的探索が必要であり、一般的な最適化手法を使うことは得策ではない。また、探索空間は、計算手法等の離散パラメータ、時間刻み幅などの連続的なものから構成され、CFL条件などの数理的拘束条件によって探索から除外される部分を持つ。さらに、ユーザが与える評価関数は簡単なものであっても強い非線形性を持つために多峰性を有する非常に複雑な探索空間になると考えられる。

このような最適化問題に対しては傾斜法よりも確率的な手法が有利である。本稿では、遺伝的アルゴリズム(GA)を利用した最適化を考える。最適化問題に対して、GAを用いる場合は、一般的には最適値のみの探索を目的とする。ここでは、ある一定世代で進化過程を止めることによって最適値を与えるスキーマの候補群を求め、それらの利用方法を考えることが目的となる。また、評価関数が複数になる場合でも多目的最適化を行えば、同じ枠組みで処理が可能である。

時間刻み幅、格子制御パラメータ、あるいは行列解法における緩和係数などの連続量は量子化によって離散パラメータ化する。または、実数型のGA¹⁵⁾を採用する。

パラメータ空間をコード化された遺伝子で表現し、評価関数から目的関数を構成して最適化問題を解く。

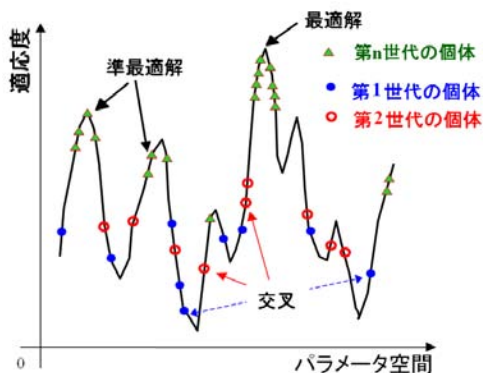


図 4 GA による最適値探索

Fig. 4 Searching for optimal value using Genetic Algorithms.

3.4 スキーマ解析によるルール抽出

パラメータ空間の複雑性から GA によっても最適解を求めることは困難である。このため、GA による探索を一定の世代で止め、その時点で残っているものから準最適解を選ぶ。初期個体を変えることによって、一連の準最適解群を保存する(図 4)。この群は探索履歴情報でもある。

これに対してスキーマ解析を行い、ルール抽出を試みる。抽出されたルールが一般化できる場合は、探索空間の制約条件として加える。このような方法によってルール抽出ができることは造船の組立工程の中で示されている⁷⁾。

求められたルールは拘束条件、あるいは個体の生成ルールとして探索プロセスにフィードバックする。

4. システムの実装

パラメータ推薦システムを実装する場合、提案したプロセスの並列性が高いことに着目すると同時に、

- 階層的、あるいは冗長な計算を許す計算環境が整い始めている、
 - ユーザの身近にはそれなりの計算資源がある、
 - 知識は個人の近くにあり広域に分散している、
- という状況をふまえる必要がある。

計算効率は、試行パラメータ全体で決まる。これは High Throughput Computing(HTC)に近い考えである。ただし、パラメータマイニングでは計算の冗長性(無駄になるであろう計算を実行する環境が存在するという考え方)を許容するということが重要である。システムを構成するうえで鍵となるのは、効果的なジョブスケジューリングである。

一方、パーソナルコンピューティング⁸⁾の普及によって数値計算に関する個々の知識が広範囲に分散するよ

うになっている。問合せシステムの役割が重要になる。

具体的な手順を示すことで、システムの実装方法を提案する。ジョブスケジューリングと問合せ部は次章以降で考察する。パラメータ推薦システムでは、問題設定と計算に対する要求項目（精度重視，計算時間重視など）を決めることから始める。その後，問題を分析し，利用可能な計算プログラムを第1パラメータ群から決定する。計算プログラムが決まれば利用できるスキームなどが決まるので探索領域を限定することができる。利用可能な機能に基づいて，設定された問題に対応する参照問題を問い合わせる。この段階で，いくつかのテスト問題，あるいは基準問題が提示される。たとえば，流体解析で移流項が重要であると判断されると，問題設定に応じて改変された移流方程式に関するテスト問題が提示される。適当な参照問題がない場合は，解くべき問題をそのまま使う。

次に評価関数と適合度を定義し，テスト問題，基準問題，あるいは与えられた問題そのものに対して最適な計算パラメータを求めるための計算を実行する。この計算は，図5に示すシステムによって行われる。

はじめに利用可能な計算機を動的に確保する（図5(A)）。次に制約条件のもとで初期個体を発生する（図5(B)）。既存の結果があるかどうかを問い合わせ（図5(C)），結果がなければ初期個体に対して計算を行う（図5(D)）。初期個体に対する評価値をもとに次世代の選択を行い（図5(E)），一定の世代までAからEを繰り返す。一定の世代となった時点で計算を打ち切り，スキーム，準最適解，あるいは振動解を保存する（図5(F)，(G)）。引き続き，新たな初期個体を発生させ，AからGを繰り返す。一定量のスキームが保存された時点でパラメータマイニングを利用してスキーム解析を行い，ルールが見つければ保存する。

5. 数値実験によるフレームワークの検証

提案するフレームワークは，計算パラメータばかりではなく，設計や物理パラメータに対しても最良パラメータが推薦できるように考えられたものである。しかしながら，すべてのパラメータを含めた場合に対してフレームワークの有効性を実証することは，問題の多様性や規模を考えると非常に難しい。CFDをはじめとする様々な分野において，モデルとなる検証問題が存在することに注目すると，実用問題のサブセットとなるモデル問題に対してパラメータマイニングを行い，その知見を実用問題に対するパラメータ推薦に活かすという戦略が考えられる。本稿では，モデル問題

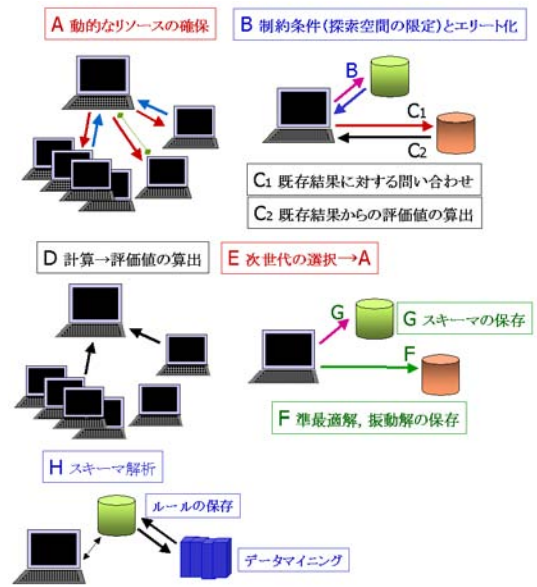


図5 システムの概念図

Fig. 5 Concept of the proposed system.

を推薦する部分は今後の課題として，それ以外の部分を検証するための実用問題での実証の第1段階と位置づける数値実験を行う。

CFDをドメインとした場合，モデル問題としては，基礎的なものと，より実用面に近いものの2種類が考えられる。前者としては，線形移流方程式，拡散方程式，バーガース方程式，スカラー場に対するポアソン方程式で表される現象に対するものが考えられる¹²⁾。後者は，キャピティ流れ，円柱を過ぎる流れ，管内流れなどの基礎的な流れに関する問題である。たとえば，高レイノルズ数流れを扱う場合は，少なくとも線形移流方程式を精度良く解くための計算パラメータが必要になる。本稿では，このような推薦部があることを前提として，基礎的なモデル問題に対して，提案するパラメータ推薦が有効であるかどうかを検証してみた。

5.1 試作システム1

推薦システムの実効性を確かめるために，PCクラスタを利用した数値実験を行う。PC間が高速回線で結ばれている必要はないので，将来動向も考慮したうえで，図6に示す無線LANによるノートパソコンを用いたクラスタシステムを構築した（表2，図6）。使用した無線LANの仕様は，IEEE802.11b（最大11Mbps）である。領域分割型の2次元非圧縮性流体解析プログラム⁹⁾を用いて並列計算性能の評価実験を行った結果は，238秒（1台：#1），1378秒（2台：#1，#2），1827秒（3台：#1～#3），2790秒（4台：#0～#3）である。そのような計算での速度向上は望



図 6 無線 LAN を用いた PC ノートクラスター
Fig. 6 PC cluster using wireless LAN technology.

表 2 ノード構成

Table 2 Configuration of the system.

node	機種名	OS	CPU	主記憶
# 0	ThinkPAD A30	RedHat7.2	1 GHz	384 MB
# 1	ThinkPAD R30	RedHat7.1	1 GHz	256 MB
# 2	ThinkPAD R30	RedHat7.1	1 GHz	128 MB
# 3	ThinkPAD A30	RedHat7.1	1 GHz	128 MB
# 4	ThinkPAD S30	RedHat7.2	600 MHz	128 MB

めないことが分かる。

パラメータ推薦システムにおけるパラメータマイニングにおける主要部は、図 7 に示すように GA を用いて準最適解となるパラメータ群を収集する部分とスキーマ解析である。通信はパラメータの付与と評価値の収集で生じるが、その量は小さい。各々の計算に要する時間はパラメータに依存して異なることから、実効性能はロードバランスのとおり方によって決まる。試作システム 1 では、GA のプログラムに計算プログラムを組み込む形で全体を構成している（並列化には MPI を用いた）。GA 部で、いくつかの初期個体を生成し、各ノードに初期個体を送り、それぞれのノードで個体をデコード後、それをパラメータとして計算が進められる。その結果は、順次 GA 部に戻される。1 世代分終了するまで、GA 部は、計算が終了したノードに対して新しい個体を送り続ける。受動的なジョブスケジューリングではあるが、この方法によって負荷の均等化を試みた。一定世代まで全体の計算を進め、その後、評価値と評価値を与えるスキーマを収集する。収集されたスキーマに対してデータマイニングを行い、ルールを抽出する。マイニングツールとして、クレメンタイン¹⁰⁾を用いた。クレメンタインとは、ニューラルネットワーク系分析ツール、決定木系分析ツール、統計解析系分析ツールを兼ね備えた汎用データマイニングツールである。

5.2 システム 1 を用いた数値実験と考察

はじめに、一次元の移流方程式に対するパラメータマイニングを例に計算効率を調べる。時間積分 {陽解法 (1 次精度, 2 次精度)}、移流項の離散化 {風上差

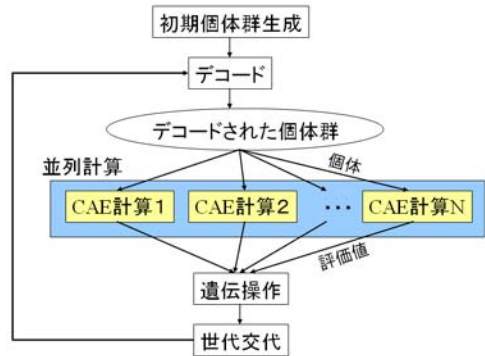


図 7 パラメータマイニングの主要部
Fig. 7 Main part of parameter mining.

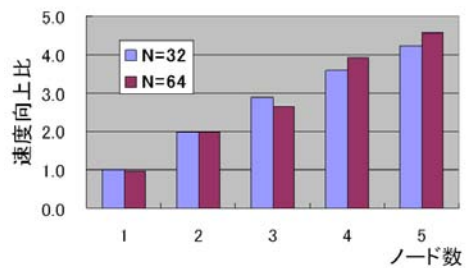


図 8 ノード数に対する速度向上比
Fig. 8 Speed-up ratio for the number of the nodes.

分 (1 次精度, 3 次精度), 中心差分 (2 次精度, 4 次精度)}, 時間刻み幅, 空間刻み幅をパラメータとし、個体数 N を変えて計算する。

1 つのジョブに要する計算時間は、パラメータによって異なる。この例題の計算時間の最長最短比は 20 倍程度であった。さらに、5 台の PC にも最大 2 倍程度の性能差がある。このような場合においても、図 8 に示すように個体数が増えたとほぼ線形な台数効果が期待できることが分かった。次に、決められた位置での厳密解との差を評価関数としたパラメータマイニングを行った。図 9 にパラメータと評価値の例を示す。100 世代後に残ったものを準最適解としてスキーマ解析を行った結果、CFL 数 1 が最も厳密解に近い解を与えるというルールが抽出できた。数値計算の知識があれば、このルールを導くことは容易である。しかしながら、このルールが既存の知見として蓄えられていない場合もあるので、数値計算の初心者に対しては、パラメータを選択する際の支援情報になりうるはずである。

さらに、振動解、発散解を与えるパラメータを負例とし、一定以上の適合度を与えるものを正例として、ニューラルネットワークを用いて教師付き学習を行い、その

パラメータ	評価値
240011100000000000	0.891
110000000100000000	1.000
110100101010101101	0.344
220000000100000000	1.000
241101010000110000	0.756
⋮	⋮

図 9 パラメータと評価値の例

Fig. 9 Example of a relation between estimated values and the parameters.

表 3 教師付き学習後の予測精度

Table 3 Accuracy of estimation after a supervised learning process.

時間	空間	Δt	Δx	予測値	実測値
2	2	0.025	0.029	0.159	0.076
1	1	0.080	0.094	0.273	0.127
1	3	0.019	0.036	0.197	0.217
2	2	0.021	0.093	0.169	0.227
1	1	0.047	0.089	0.270	0.329
2	4	0.014	0.017	0.422	0.398
2	4	0.062	0.022	0.529	0.538
2	4	0.023	0.014	0.687	0.711
1	1	0.100	0.010	0.902	1.000
2	2	0.100	0.010	0.899	1.000

学習結果を保存した。ある程度、学習が進んだ段階で、ユーザが適当なパラメータを与えると、表 3 に示すように高い精度で評価値が予測できることが分かった。ここで、時間の欄の 1, 2 は、1 次精度, 2 次精度の時間積分を、空間の欄の 1, 2, 3, 4 は、1 次風上, 2 次中心, 3 次風上, 4 次中心差分を示している。

解くべき問題の性質が異なれば、このような教師付き学習の利用は難しくなる。しかしながら、データ生成プロセスの類似性を考慮し、非線形性に注意すれば、同じような傾向を示すものを提示することができるはずである。この場合、通常は棄却されるデータを系統的に蓄積することが重要である。最後に、二次元移流方程式に対するパラメータマイニングを行った結果、図 10 に示すいくつかのクラスタが形成された。これは、移流項の離散化、格子刻み幅、時間刻み幅に対して、数値安定性、あるいは CFL 数では解釈できない何かしらのルールの存在を示唆する結果である。

これらは単純な例題から導かれた結果であり、提案した方法によってただちに未知のルールが導けると結論づけることはできない。しかしながら、単純な問題であるにもかかわらず、解析的には分かっている数値解の挙動が、実際の計算を行ううえで高精度化や効

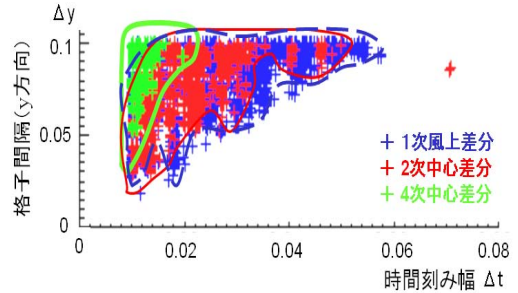


図 10 ある適合度以上を与えるパラメータの散布図

Fig. 10 Scatter diagrams of the parameters produce higher goodness-of-fit values.

率化に寄与する。解析的アプローチは重要であるが、本稿で提案する経験的方法からのルール抽出の可能性を見逃してはならないだろう。この可能性を現実化するためには膨大な数のパラメータスタディが必要になる。1 つ 1 つの計算時間はそれほど大きいものではなく、通信量も小さい。これをふまえてシステム 2 を構築する。

5.3 試作システム 2

試作システム 1 の問題点は、PC クラスタをベースとした専用システムを用いていること、同一プログラム中で GA 部と計算部を制御していることである。

試作したシステム 2 は、Grid 計算環境を模擬したシステムである。構成する計算機としては、表 2 で示すものにデータベースサーバとして #5 (Pentium4 1.8 GHz, 主記憶 512 MB, Red Hat 7.3) を加えたものである。Globus Toolkit¹³⁾ をすべてのノードに入れ、#0 を GRAM サーバとした。ローカルの資源管理ツールとして、OpenPBS¹⁴⁾ を利用している。

GA 部と計算部は分離し、GA のプログラムと計算プログラムは、ファイル I/O によってやりとりをする。このようにして既存の計算プログラムを推薦システムに組み入れることが容易になる。

システムの構成は、図 5 に示したものに準拠している。各計算機においてジョブが終了すると、ジョブの終了と (パラメータ, 評価値, 計算データ) の格納場所が URL としてクライアント側に伝えられる。1 世代分終了するまで、ジョブが終了している計算機に対して、デコードされた個体 (パラメータ) が送られ、計算実行を促す点は、システム 1 と同じである。1 世代終了後、遺伝操作が行われ、新たな個体群に対して一定世代まで計算が繰り返される。知識抽出の部分は、システム 1 と同様であるが、システム 2 では、ある程度適合度の高い (パラメータ, 評価値, 計算データ) をデータベースとして蓄えることによって、データの

再利用性を高めている。データ生成プロセスに関しては、計算パラメータとともに計算内容を XML で記述し¹¹⁾、データの一元管理を行う。現時点では、既存データの検索はパラメータを検索キーとして行い、検索結果を URL で返している。

システムの計算効率を確かめるために、一次元移流方程式に対する計算プログラムを用いて、計算パラメータを変えた 101 個のジョブを実行してみた。はじめに、#0 と同じ仕様の PC を用いて計算時間を計測した。各ジョブの計算時間にはバラツキがあり、最長と最短の比は約 30 であった。次に、全体の計算時間をもとに台数効果を調べると、#1~#4 を利用した場合には、ほぼ線形の上昇比となることが分かった。ただし、#0 を含めた速度向上比は 3 倍程度であった。#0 は計算ノードではあるが、GRAM サーバでもあり、ジョブの送信および起動のプロセスに計算時間が費やされて全体の効率が落ちるためである。この効率劣化は、プロセス管理、ジョブスケジューリングの工夫によって改善できると考えられるが、今後の課題である。

次に、計算結果群が存在する場合について計算効率を調べた。データベースサーバ(#5)を加え、はじめに任意に選んだ 30 個のジョブを #0~#4 で計算した。計算データをそれぞれのノードに残し、URL で示された計算データのありか、そのパラメータをデータベースサーバで管理する。その後、101 個のジョブを、#0~#4 で実行した結果、計算時間は約 23% 減少した。

システム 2 は広域分散計算に対する適応性が高く、Grid のアプリケーションとしても有効であると考えられる。

6. 結 語

計算パラメータの推薦システムを構築するために、パラメータスタディにおける課題を明らかにし、システムに対する要件をまとめ、システムの概念設計を行った。

システムには、評価値を獲得・管理するための効率的な方法、評価値の分析から次のパラメータスタディに至る連続したプロセス、プロセスからの知識抽出方法が必要であることを示した。

提案したパラメータ推薦システムのフレームワークをまとめれば、

- 評価値の効率的取得・管理、知識抽出のために、データ生成プロセスをパラメータ化すること、また、パラメータを階層化すること、

- 最良パラメータの探索をユーザが設定した評価値に対する最適化問題に結び付けること、
- 最適化問題を遺伝的アルゴリズムによって解く過程で準最適解群を求めること、
- スキーマ解析によってルール抽出を行うこと、
- ルールを次のパラメータ探索に活用するというフィードバックプロセスを形成すること、
- 大量のパラメータ検索実行を効率的に行うための仕組みを作ること、

になる。

フレームワークの実効性を確かめるために、2 つの試作システムを作り、数値実験を行った。その結果、パラメータ推薦システムの基本概念の有用性が確認できた。また、システムにおける計算部の並列性は高く、通信量が少ないために粗結合の並列環境であっても効率的なパラメータ推薦が可能であること、データの蓄積・管理を含めて広域分散計算に適したアプリケーションであることを示せた。

CAE の適用問題が拡大するとともに、CAE の専門家以外の利用がさらに増えるものと予想される。そのような利用者を支援するシステムが CAE の普及にとって重要である。実用問題への適用、効果的なジョブスケジューリングなど今後の課題も多いが、提案した計算パラメータ推薦システムの概念や手続きは、他のパラメータに対しても適用できるものであり、結果の解釈に至る工程を短縮できるものであると考えられる。

参 考 文 献

- 1) Shirayama, S. and Kuwahara, K.: Navier-Stokes solution of the flow field around a complete automobile configuration, *Proc. 2nd International Conference on Supercomputing in the Automotive Industry*, pp.293-304 (1988).
- 2) Shirayama, S.: Local network method for incompressible Navier-Stokes equations, AIAA-91-1563-CP (1991).
- 3) 柄谷和輝, 中村 壽, 奥田洋司, 矢川元基: 並列有限要素法コード GeoFEM の性能評価, 日本計算工学会論文集第 2 巻, 20000022, pp.171-177 (2000).
- 4) 太田高志, 白山 晋: オブジェクト指向フレームワークによる流体計算統合環境, 日本計算工学会論文集第 1 巻, 19990001, pp.27-33 (1999).
- 5) ICEM CFD Engineering.
<http://www.icemcfd.com/>
- 6) SMARTFIRE.
<http://fseg.gre.ac.uk/smartfire/home.html>
- 7) 大和裕幸, 白山 晋, 増田 宏, 佐野 毅: 遺

伝アルゴリズム手法による造船組立工程のルール抽出に関する研究, 日本造船学会論文集第 188 巻, pp.559-568 (2000).

- 8) 白山 晋: 科学技術計算におけるパーソナルコンピュータの可能性について, パーソナルコンピュータユーザ利用技術協会論文誌, Vol.8, No.1, pp.31-40 (1999).
- 9) http://www.nakl.t.u-tokyo.ac.jp/~sirayama/index_tb.htm
- 10) クレメンティン .
<http://www.spss.com/spssbi/clementine/>
- 11) 白山 晋, 大和裕幸: 可視化情報分析支援システムのフレームワークについて, 日本造船学会論文集第 190 巻, pp.267-275 (2002).
- 12) たとえば, 棚橋隆彦: はじめての CFD—移流拡散方程式, コロナ社 (1996).
- 13) Globus Toolkit. <http://www.globus.org/>
- 14) OpenPBS. <http://www.openpbs.org/>
- 15) たとえば, Michalewics, Z.: *Genetic Algorithm + Data Structures = Evolution Programs*, Springer-Verlag (1992).
- 16) Berners-Lee, T.: *Weaving the Web*, Orion Business Books (1999).

(平成 14 年 9 月 23 日受付)

(平成 15 年 1 月 4 日採録)



白山 晋

昭和 34 年生. 昭和 57 年京都大学工学部航空工学科卒業. 昭和 62 年東京大学大学院工学系航空学専攻博士課程修了 (株) 計算流体力学研究

所 (株) ソフテックにおいて, 流体解析手法, 流体解析システム, 可視化手法, および可視化システムの研究に従事. 平成 11 年より東京大学大学院工学系研究科助教授. 現在東京大学人工物工学研究センター助教授. 大規模データマネージメント, セマンティックグリッド, 可視化情報からの知識抽出に関する研究に従事. 工学博士. 平成 14 年可視化情報学会論文賞受賞. 可視化情報学会, 日本計算工学会, 日本造船学会, 日本機械学会, 日本流体力学会各会員.



齋藤幸二郎

昭和 50 年生. 平成 13 年東京大学工学部船舶海洋工学科卒業. 平成 15 年東京大学大学院工学系研究科環境海洋工学専攻修士課程修了. CFD における可視化, グリッドコンピューティングの研究に従事. 平成 15 年 4 月より日本アイ・ビー・エムサービス事業部, 金融第一ソリューション・センター勤務. 工学修士. 平成 14 年日経サイエンスビジュアルサイエンスフェスタ佳作. 日本造船学会会員.



竹森 恵一

昭和 55 年生. 平成 14 年東京大学工学部システム創成学科卒業. PC クラスタによる流体解析の研究に従事. 平成 14 年より東京大学大学院工学系研究科環境海洋工学専攻修士課程在学中. オホーツク海における氷況シミュレーション手法の開発の研究に従事.



太田 高志 (正会員)

昭和 36 年生. 昭和 60 年慶應義塾大学理工学部物理学科卒業. 昭和 62 年慶應義塾大学大学院物理学専攻修了. 平成 2 年東京大学大学院航空学専攻単位取得退学. 平成 3 年より日

本アイ・ビー・エム東京基礎研究所勤務. 物理シミュレーション, 並列・分散計算, グリッドコンピューティングの研究に従事. 工学博士. 平成 14 年可視化情報学会論文賞受賞.