

平安京ビューによるシミュレーションパラメータの対話的最適化

伊藤 貴之[○], 小山田 耕二 皿井 伸明 比戸 将平 (京都大学)

Interactive Optimization of simulation parameters by Heiankyo-View

Takayuki ITOH Koji KOYAMADA Nobuaki SARAI Shohei HIDO

ABSTRACT

Computational simulation became popular not only in the fields of physics or chemistry, but also medical, engineering, finance, and so on. Discovery of the best simulation parameters is an issue in order to fit the result of simulations to real experimental results. This paper presents a system that realizes interactive optimization of such simulation parameters with a visualization technique and distributed computing architecture. The system iteratively executes simulations with gradually varying parameters, and then records the errors of simulation results. Then the system visualizes the distribution of the errors by Heiankyo-View, a novel hierarchical data visualization technique. We assume that there are multiple local-minimum errors in the distribution, but the visualization results help users to interactively choice the best parameters, or continue simulations with interactively varying the parameters.

Keywords: Visualization, Heiankyo-View, Interactive optimization

1. はじめに

計算機シミュレーション技術は近年、理学、工学、地学、医学、経済学、など多くの分野に広がりを見せている。計算機シミュレーション技術のコアとなる数式モデルには、多くのパラメータを有する 경우가多く、その適切な値を発見することは非常に重要である。計算機シミュレーションを実用する多くの現場では、まずシミュレーションの対象における実測値を入手し、その実測値に算出結果が近づくように計算機シミュレーションの数式モデルやパラメータ値を選択することで、計算機シミュレーションを信頼できるものにする。これが確立して初めて、計算機シミュレーション技術は多くのデータ・多くの現象に広く適用することが可能となる。

ここで、計算機シミュレーションの n 個のパラメータを $x_1 \sim x_n$ とし、シミュレーション結果の実測値に対する誤差を $f(x_1, \dots, x_n)$ とすると、最適なパラメータを求めるとい問題は f 値が十分小さくなるようなパラメータ値を求めるとい問題に帰着される。しかし実測値自体の誤差などを考えると、必ずしも f 値が最小になるパラメータが最適なパラメータであるとも言い切れない。むしろ、 f 値がある程度小さくなる何組かのパラメータ値を提示し、知識や経験則を有するシミュレーション熟達者がその中から対話的に最適なパラメータを選択する、というようにシステムを構築することが望ましい場合もある。そこで、パラメータ値の変化に対するシミュレーション結果の誤差分布を視覚技術と GUI によって提示し、熟達者が対話的に最適なパラメータを発見できるシステムを構築する、という点が本研究の目的である。

ここで、例えばパラメータが1個や2個で、シミュレーション結果が数個~数十個、というような小規模な数値分布であれば、日常的に用いられている表計算ソフトウェアを用いることで、その数値分布をある程度理解することが可能である(図1参照)。しかし、パラメータが5個、10個...と増大したり、あるいはシミュレーション結果が数百個、数千個...と増大したりする場合には、図1に示すような表示技術でシミュレーション結果の誤差分布を理解できるとは言えない。

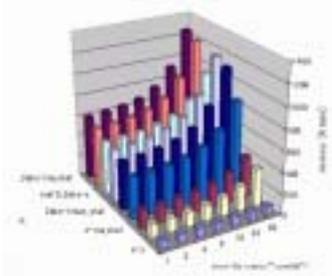


図1 表計算ソフトが提供する3次元棒グラフ表示機能。

本報告では、大規模階層型データの視覚化手法「平安京ビュー」[Ito04]を用いて、多数のパラメータから得られるシミュレーション結果の誤差分布を視覚化する手法と、それを用いたシミュレーションのための分散環境を提案する。この環境を用いて対話的にパラメータを探索しながらシミュレーションを反復することで、ユーザはシミュレーションの最適なパラメータ群を発見できると考えられる。

2. 平安京ビュー

前節でも述べたとおり、「平安京ビュー」は大規模階層型データの視覚化手法である。図2に「平安京ビュー」による階層型データの表示例を示す。「平安京ビュー」では、階層型データを構成する葉ノードをアイコンで、枝ノードを長方形の枠で表現し、長方形枠の入れ子構造によって階層構造を表現する。

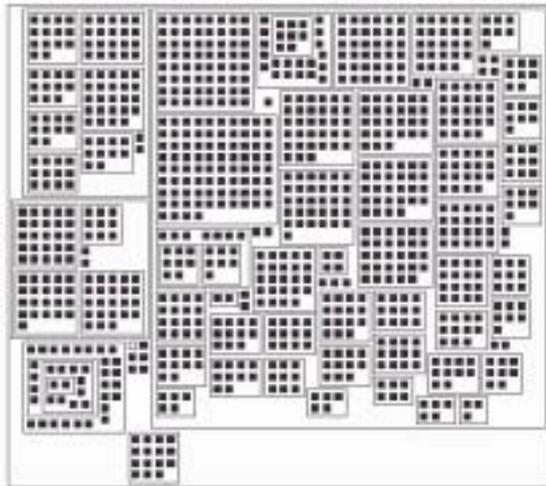


図2. 平安京ビューによる階層型データの表示例。

「平安京ビュー」では、まず最下位階層に属する葉ノードに対応するアイコンを、画面空間上に隙間無く配置する。続いて、この上位階層に属する枝ノードを表現するために、アイコンを包括する長方形を生成する。さらに、上位階層の枝ノードを表現する長方形群を隙間無く配置し、同様にこれを包括する長方形を生成する。以上の処理を、最下位階層から最上位階層に向けて反復することで、データ全体の配置を決定する。図3に、「平安京ビュー」による葉ノードおよび枝ノードの画面配置手順を示す。

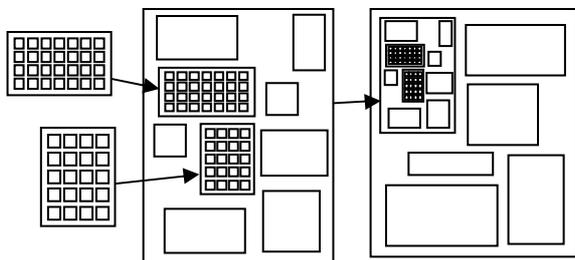


図3 階層型データの画面配置順。まず最下位階層の葉ノードを配置し、続いて下位階層から上位階層に向かって配置処理を反復する。

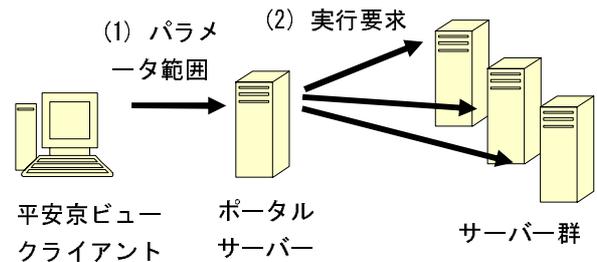
このとき「平安京ビュー」では、葉ノードおよび枝ノードを表現する長方形群を、以下の2条件を満たすように画面配置する。

[条件 1] すでに配置されている長方形と干渉しない位置に長方形を配置する。

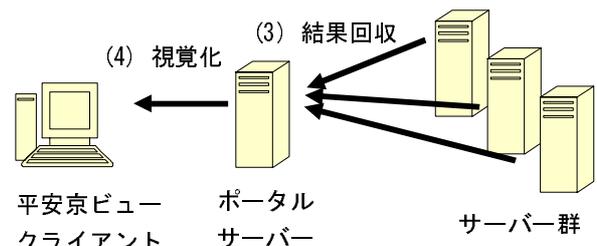
[条件 2] [条件 1]を満たす位置のうち、配置面積の拡大量が最小である位置に長方形を配置する。

この条件を満たすような画面配置を実現することによって、大規模な階層型データの全貌を、限られた画面空間にあますことなく表現することができる。

3. シミュレーション結果の誤差分布を提示するアーキテクチャ



(a) シミュレーションの実行要求



(b) シミュレーション結果分布の視覚化

図4 シミュレーション結果分布の視覚化のアーキテクチャ。

図4に、本報告が提案するアーキテクチャの概要図を示す。このアーキテクチャは、以下のような処理手順を反復することを想定している。

- (1) クライアントからポータルサーバーに対して、各々のパラメータ x_i に対して最小値と最大値で表現される区間 $[x_{imax}, x_{imin}]$ を与える。 n 個のパラメータを有するシミュレーションの場合、 n 個の区間 $[x_{1min}, x_{1max}] \dots [x_{nmin}, x_{nmax}]$ を与える。
- (2) ポータルサーバーからサーバー群に対して、パラメータの値を変化させて渡すことで、シミュレーションの実行を要求する。 n 個のパラメータ範囲をそれぞれ m 区間に分割した場合、シミュレーションの合計実行回数は m^n となる。
- (3) サーバー群によるシミュレーション結果を、ポータルサーバーが集める。ポータルサーバーはシミュレーション結果と実測値との誤差を算出して記録する。
- (4) クライアントがポータルサーバーにシミュレーション結果を要求すると、ポータルサーバーはシミュレーション結果の誤差値の一覧を返す。クライアント

は「平安京ビュー」を用いてシミュレーション結果の誤差分布を視覚化する。

- (5) 「平安京ビュー」の画面をみながら、新しいパラメータ区間を対話的に指定し、(1)に戻ってシミュレーションの要求を反復する。

なお本報告が提案するアーキテクチャでは、サーバーで稼動するシミュレーションの所有者と、シミュレーションの利用者が、別の人である場合などを想定している。このとき、クライアントはサーバーに対して LAN 接続できない場合も考えられる。本報告では、インターネット上で計算機アプリケーションを自在に接続する Web サービスという計算機基盤技術 [WS] を、クライアント～ポータルサーバー間の通信に用いることを想定する。Web サービスでは、サーバー～クライアント間のインタフェースを公開することや、またその間で交換するデータやプロトコルを XML 技術によって共有することが容易であり、特に異組織にまたがる分散計算環境に適していると考えられる。

また本報告が提案するアーキテクチャでは、シミュレーションの計算負荷が非常に高いこと、あるいはシミュレーションの回数が非常に多いこと、なども想定している。このような場合において、計算機資源の有効活用は非常に重要である。計算機資源を有効活用するための一手段として、ポータルサーバーが自在にサーバー群の運用状況を監視し、最適な計算機を動的選択してシミュレーションを実行要求することが考えられる。このようなメカニズムを実現するために、本報告では、ポータルサーバーとサーバー群の運用にグリッドコンピューティング [Grid] を適用することを想定する。

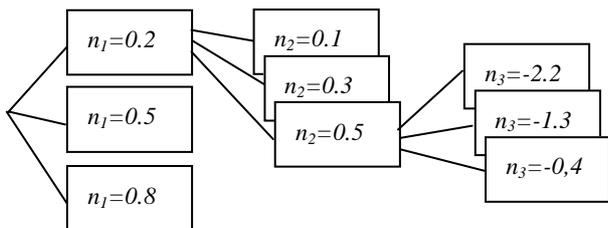


図5 パラメータ値からの階層型データの構築。

ところで、「平安京ビュー」は階層型データを対象とした視覚化手法である。よってシミュレーション結果の誤差分布の視覚化に「平安京ビュー」を適用するには、シミュレーション結果の誤差値群を階層型データに変換する必要がある。本報告では単純に、 n 個のパラメータの中から 1 個を選んで、そのパラメータ値でシミュレーション結果の誤差値群を分類し、続いて別のパラメータを 1 個を選んで、そのパラメータ値でシミュレーション結果の誤差値群を分類し... という操作を n 回反復することで、階層型データを構築している (図 5 参照)。

4. 細胞シミュレーションへの適用

筆者らは現在、本報告で提案するアーキテクチャを、細胞シミュレーション結果の誤差分布表示と、その対話的なパラメータ最適化の目的で適用している途中である。本章では、その細胞シミュレーションについて紹介する。

生命工学の分野では近年、情報処理技術を用いて生命のメカニズムを解明しよう、という研究が急速に発達している。特に人間の遺伝子 (いわゆるヒトゲノム) の規則性を計算機によって読解するという研究は、ここ数年間の急速な研究により、ほぼ終息したと言える。しかし遺伝子を読むだけでは、生命のメカニズムには解明しきれない。そこで別のアプローチとして、生体を構成する単一機能要素を微分方程式系のモデルとして表現し、その複合系として細胞モデルや組織モデルを構成し、情報処理技術によってそのメカニズムを観察する、という研究が進んでいる。

著者らが適用する細胞シミュレーション手法 "simBio" [Sar04a][Sar04b] では、機能を表現する微分方程式などの数式モデルを Reactor, それに対するパラメータを Node として、Java を開発言語としたオブジェクト指向プログラミングにより、生体機能の計算機シミュレーションを実現している。図 6 に機能モデル構築の一例を示す。図 6 において、Reactor は長方形で、Node は円で記述されている。このように simBio では、少数の数式モデルと多数のパラメータをオブジェクトとして実装し、そのグラフ構造として機能モデルを構築し、計算機シミュレーションによってその挙動を観察している。

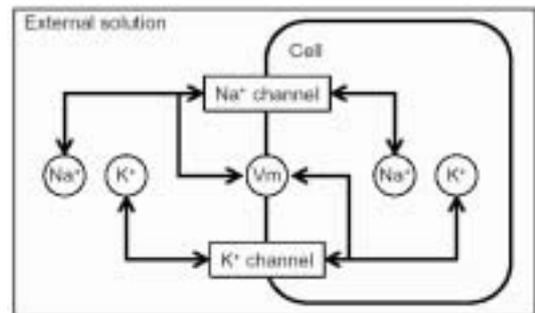


図6 Reactor と Node による生体機能モデルの構築

simBioを用いて心室筋細胞を複合系モデルとして表現し、計算機シミュレーションによってその時間応答を得た結果の例を図7に示す。図7の左側の表は、このシミュレーションに適用した膨大な数のパラメータ(Node)であり、右側の折れ線グラフは、活動電位や、各イオンの電流量、などの時間応答を示している。これらの時間応答に対して実測値を持っていれば、この実測値との誤差を算出することにより、計算機シミュレーション結果の妥当性を検証することができる。

simBioでは、計算機シミュレーションの入力値であるパラメータ群、および出力値である時間応答結果に対して、XMLを用いた記述書式を定義している。またそれ以外の

機能についても、Java を用いてプラットフォーム非異存なシステム構築を実現している。これは simBio が、多数の研究者および組織にまたがって標準化されたデータ記述技術とインタフェースによって計算機シミュレーション環境を共有し、オブジェクト指向プログラミングによって数式モデルとパラメータの差し替えだけで多彩な実験が進められるように、という理念に基づいて設計されたことに起因する。この理念はまさしく、3 章で示したアーキテクチャの理念に合致しているといえる。

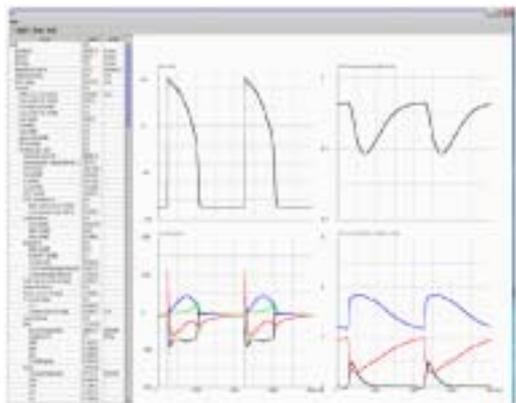


図 7 simBio を用いた心室筋細胞のシミュレーション

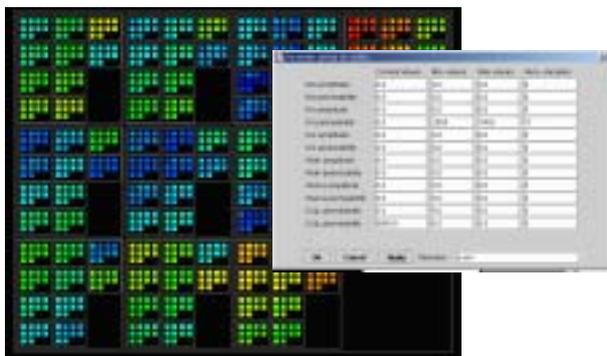


図 8 simBio による計算機シミュレーション結果の誤差分布の視覚化結果と、そのパラメータ再入力のための GUI

図 8 に、simBio を適用した計算機シミュレーション結果の誤差分布の表示結果と、その結果を参照しながらパラメータを再入力できる GUI の例を示す。この例では、simBio の入力値のうち 3 種類のパラメータを 10 段階の可変値とし、その組み合わせを用いて $10^3=1000$ 回の計算機シミュレーションを実行している。そして、その結果として得られる 1000 個の誤差値を、階層型データに格納して視覚化している。さらに、視覚化結果のうちの特定部位をクリックし、その部位に対応するパラメータ値を固定値として、他のパラメータを可変値として GUI 上で指定することにより、さらに好ましいパラメータを探索するために計算機シミュレーションを再実行できる。

5. おわりに

本報告では、「平安京ビュー」を用いてシミュレーション結果の誤差値の分布を視覚化する手法と、これを用いて対話的に最適なパラメータ値を探索するためのアーキテクチャを提案した。

今後の課題として、以下のような技術的拡張を検討している。

3 章で述べたとおり本手法では、 n 個のパラメータを m 区間に分割した場合、実行要求するシミュレーションの合計実行回数は m^n となる。よって、パラメータ数が非常に多いシミュレーションの場合、シミュレーションの実行回数が膨大な数になることが考えられる。この問題を解消する一案として、少ないシミュレーション結果から、その誤差値を補間することで、最適なパラメータ値を類推することが考えられる。このような手法と「平安京ビュー」との組み合わせで、さらに効率的に最適なパラメータを探索することができると考えられる。

また、本手法が示すアーキテクチャは、シミュレーション実行と視覚化を非同期なプロセスとして実装可能である。よって、シミュレーションの進行途中段階であっても、その途中段階を対象とした誤差分布の視覚化が可能である。よってシミュレーションの結果分布を視覚化するだけでなく、計算途中のシミュレーションの進行状況を監視する目的でも用いることができる。並列計算環境の稼働状況を監視する目的での視覚化手法は過去にも報告されており [Yam03]、その目的とシミュレーション結果分布の視覚化との両方の目的を同時に達成する手法が確立できれば、より有用性が高まるものと考えられる。

謝辞

本報告のアーキテクチャに関して京都大学細胞生体シミュレーションプロジェクト関係各者から貴重な意見を賜ったことを感謝いたします。「平安京ビュー」の利用方法に関して日本アイ・ビー・エム (株) 東京基礎研究所山口裕美氏から貴重な意見を賜ったことを感謝します。

参考文献

- [Ito04] 伊藤, 山口, 小山田, 画面空間の格子分割を用いた階層型データ視覚化手法, 情報処理学会論文誌, 投稿中.
- [WS] Web Services Activity, <http://www.w3.org/2002/ws/>
- [Grid] Globus Project, <http://www.globus.org/>
- [Sar04a] 皿井, 天野, 松岡, 松田, 野間, 生物学的視点に基づくオブジェクト指向生体機能シミュレーション, シミュレーション, Vol. 23, No. 1, 2004.
- [Sar04b] 皿井, 野間, simBio: 生物学的ダイナミックモデル開発基盤, 生体医工学誌, 2004 年 2 月
- [Yam03] Yamaguchi Y., Itoh T., Visualization of Distributed Processes Using “Data Jewelry Box II” Algorithm, CG International 2003, pp. 162-169, 2003.