

網羅的なシミュレーションにおける 空間的パターン発見のためのデータマイニング手法の開発

石田 愛[†] 渡辺 知恵美[‡] 城 和貴[†]

[†] 奈良女子大学人間文化研究科 〒630-8506 奈良県奈良市北魚屋西町

[‡] お茶の水女子大学理学部 〒112-8610 東京都文京区大塚 2-1-1

E-mail: [†] {ishida, joe}@ics.nara-wu.ac.jp, [‡] chiemi@is.ocha.ac.jp

あらまし 近年, コンピュータ技術の進歩により様々な要素が複雑に組み合わさっている物理シミュレーションは常識となってきた。技術者らは, 膨大なデータの中から複数の要素の分布における空間関係や, 複数のパラメタ間における物理現象の相関関係発見の手がかりを発見した後, 詳細に分析するが, その手がかりの発見は技術者の経験的知識に基づいた直感によることが多い。そこで, 我々は, 物理シミュレーション結果の相関関係に関する手がかりの発見を支援するために, 考えられる全ての条件を網羅し試行した結果の鳥瞰図を提示するマイニング手法を提案する。これにより, 直感的にどの領域でどのような特徴が出ているのか大まかに知ることができ, 調査する領域を絞り込むことが可能になる。

キーワード 科学 DB, データマイニング, 可視化, 時空間 DB

A Data Mining Method for Finding a Spatial Correlation Between Two Volume Datasets

Ai ISHIDA[†], Chiemi WATANABE[‡] and Kazuki JOE[†]

[†] Graduate School of Humanity and Science, Nara Women's University Kitauoya nishi-machi, Nara, 630-8506, Japan

[‡] Faculty of Science, Ochanomizu University 2-1-1, Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan

E-mail: [†] {ishida, joe}@ics.nara-wu.ac.jp, [‡] chiemi@is.ocha.ac.jp

Abstract In recent years, with the advances in computer technology, computer simulation program can calculate the various factors which are related to the phenomena, and it is important to analyze the correlation of these factors. For efficient analysis of correlations between some simulation data, engineers often look for some clues of correlation feature to analyze in detail. But the way of finding clues depends on an engineer's heuristics and technical knowledge. In this paper we propose a visual mining technique that supports to find some clues of characteristic correlations of two simulation data. We suppose that the simulation data is formatted as 3d-gridded volume data. In the mining technique, the regions are extracted by multiple conditions such as "200k < temperature < 220k" from each data. Next we make the "correlation-map", which is a matrix and each cell has the feature value which represents spatial relationship between regions which are selected from each data, and we visualize the matrix by using color-space model. By using this technique, engineers can find roughly where a correlation feature appears and reduce the time one takes to analysis.

Key words scientific DB, data mining, visualization, spatio-temporal database DB

1. はじめに

近年のコンピュータ技術の進歩により, 工業製品の製造過程においても物理シミュレーションを取り入れた設計プロセスが常識となっている。工業製品などの物理シミュレーションでは CAD システム等で設計された幾何形状に対して,

構造, 流体, 電熱, 振動, 騒音などのシミュレーションを行い, それぞれを可視化し比較検討する。一般に物理現象は温度や圧力, 速度場など様々な要素が複雑に絡み合っていることから単一の要素の分布や流れを調べるだけでなく, 複数の要素の分布や動きの相関関係を検証する必要がある。

る。

例えば自動車のエンジン内の燃焼において、温度と窒素濃度に何かしらの相関があるかもしれないと考えたとき、技術者は相関関係を検証する初期段階として、可視化システムなどを用いて、ある範囲の温度領域とある範囲の窒素濃度を等値面等で表示して空間的な相関関係を視覚的に確認する場合がある。その際、技術者は経験的知識に基づいた直感により試行錯誤的に様々な条件で関連のありそうな複数の要素の領域を比較し、そこから何かしらのヒントを得ようとするだろう。

また一つの物理シミュレーションを行う際にも複数の計算パラメータを試行錯誤的に変更して計算を繰り返し、それらのパラメータ間における物理現象の相関関係を検証したい場合がある。図1は2つの正方形の柱をくいちがいに配置した場合の流体の流れをシミュレーションしたものである[1]。図1(a)で鉛直方向における柱の位置関係が比較的近いが、図1(b)では柱の距離を鉛直方向に少し広げたため、柱の間に流れが発生している様子が分かる。垂直方向の距離を a 、水平方向の距離を b と置いたとき、それらのパラメータを変化させることにより、様々な流体のフローパターンを確認することができる。

例に挙げたような試行錯誤は比較的短時間な計算で気軽に行うことができるとする。その場合、可視化システムなどを用いて対話的に比較検討をする方法もある。しかしながらパラメータ値を少しずつ変えながら表示を切り替えて検証をするうちに、表示した全ての結果を1枚の紙の上にならりと並べて、パラメータを変えることによって状態がどのように変化したかを鳥瞰的に視認したいと思うこともあるだろう。図1(c)では垂直方向の距離 a 、水平方向の距離 b という二つのパラメータ値による柱周りの流体シミュレーションを網羅的に検証し、その流体パターン2次元図で表示し、さらにそのパターンを7種類に分類している。この図に示されるように、全てのパラメータによる計算結果を並べてその全容を鳥瞰し、パラメータの値による計算結果の遷移を発見する手がかりとする手法も有用であると考えられる。

そこで我々は、比較的安価に計算が実行できる物理シミュレーションにおいて、複数の計算結果データの相関関係や、複数のパラメータ間による計算結果の遷移状況を発見するための手がかりを得るために、考えられる全ての条件を網羅し試行した結果の鳥瞰図を提示するビジュアルマイニング手法を提案し実装する。具体的には、2種類のパラメータ値 i, j によって得られるデータセット $D(i, j)$ に対して、そのデータセットの特徴値 $R(i, j)$ を求める。そしてその特徴値を色空間にマッピングし、2種類のパラメータ値 i, j を軸に持つ2次元図の上にタイル上に配色する。これにより、ユーザはパラメータ値 i, j によるデータセット $D(i, j)$ の遷移を配色の遷移によって鳥瞰的に確認することができる。

本提案手法を実現するにあたって考慮しなければならな

いことは、データセットにおける特徴値は対象データや目的によって大きく異なるという点である。そのため、各対象データによって特徴値をカスタマイズしなければならず、汎用性が狭まってしまう。そこで、我々は2つの3次元ボリュームデータを対象とする。2つのボリュームデータ A, B における値 v_a および v_b をパラメータ値とし、各々の値によって特定される空間領域(等値面) $S(v_{ai})$ および $S(v_{bj})$ における空間関係 $R(S(v_{ai}), S(v_{bj}))$ を特徴値としてビジュアルマイニングを行う。二つのボリュームデータにおいて等値面を比較することは、データの種類や目的に限らず可視化による分析の初歩段階で行われることであり、われわれが行うマイニングの結果も分野を問わず有用になると思われる。

本稿では、まず第2節で提案手法の概要を述べる。第3節でそれらバリエーションについて述べる。そしてそれらバリエーションの中から汎用的に使用できるものを第4節において検討する。

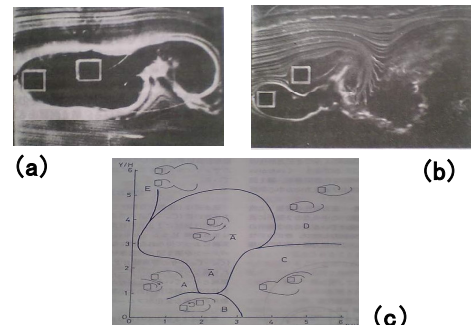


図1：流体の流れのシミュレーション

2. 提案手法の概要

本稿で提案するのは、例えば温度や湿度、濃度分布などのデータがもつ複数の要素の分布における空間的関係や複数のパラメータ間における物理現象の相関関係の手がかりを発見するために、網羅的な試行結果の鳥瞰図を提示するマイニング手法である。図2では自動車のエンジン内の燃焼における、ある温度領域と窒素濃度領域との相関関係を検証する例に対して本提案を適用した際の処理のフローイメージを示している。まず、窒素濃度が i 以上 $i + \Delta d_i$ 未満の条件を満たす空間領域 $S(i)$ と、温度が j 以上 $j + \Delta d_j$ 未満の条件を満たす空間領域 $S(j)$ を求める(図2(a))。次に領域 $S(i)$ と $S(j)$ による特徴ベクトル $R(S(i), S(j))$ を求め(図2(b))、この特徴ベクトルの値を x 軸に窒素濃度、 y 軸に温度を割り当てた格子のセル (i, j) 上に割り当てる(図2(c))。以上の処理を格子状のそれぞれのセルの x 値および y 値の範囲に対応する空間領域全てに対して繰り返し行い、窒素濃度-温度の格子状空間のセルを全て特徴ベクトルで埋めていき、最後にこれらの特徴ベクトルを色に割り当てて表示させる(図2(d))。このようにして窒素濃度および温度の取りえる範囲の全ての条件で網羅的に空間関係を調べ、格子状にカラーマッピングした結果、図2の(d)に見られるようなタイ

ル上の画像が出力される。この画像を温度と窒素濃度の空間的相関関係のオーバービューとみなし、ビジュアルマイニングを試みる。ビジュアルマイニングとは、人の視覚を利用することによって、どのあたりで特徴的な現象がおきているか、もしくはどの条件範囲とどの条件範囲に分類できるかなどを判別する手がかりとするものである。

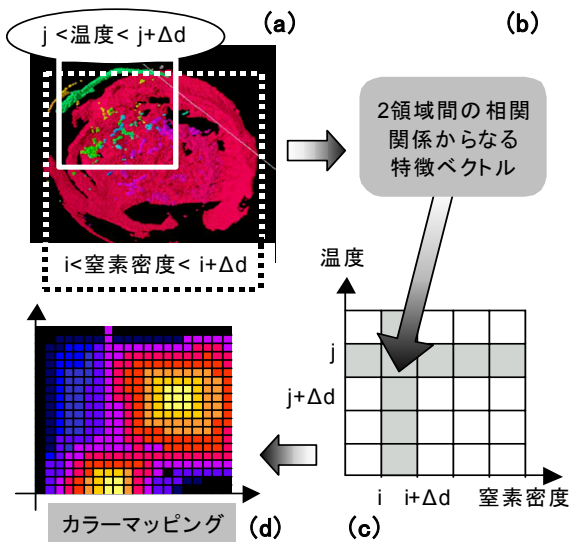


図 2：提案するマイニング手法のイメージ

2.1. 空間的相関関係を表現する特徴ベクトルの生成

本論文では網羅的に行われたシミュレーションデータ間の相関関係による空間的特徴を抽出するためのマイニング手法を提案するが、その前にまずはシミュレーションデータから特徴ベクトルを求める必要がある。例えば前節でも述べた自動車のエンジン内の燃焼における温度と窒素濃度の相関を調べたい場合、まずは温度を x 、窒素濃度を y として 2 次元格子を作成し、格子の各セルは温度 x_n の空間領域と窒素濃度 y_m の空間領域との空間関係とする。空間関係として考えられる基準としては、まず Open Geospatial Consortium(OGC)モデルにて定められている a)距離関係 b)方向関係 c)位相関係があげられる[2]。特に位相関係は 9-intersection モデルにより、二つの領域が同じ (equal)、一方が他方に含まれている (inside)、離れている(disjoint)などの二つの空間の位置関係や方向関係等を示すものである 8 種類の位相関係に分類されており、これらを 0~7 の値に割り当てることによって 1 次元の特徴値として表現できる [3]。そしてこれらの空間的関係を複合的に利用し多次元の特徴ベクトルで表現する。以上のように全てのセルの特徴ベクトルを求め、それによって作成された格子に対してビジュアルマイニングを行っていく。

2.2. カラーマッピングによるビジュアルマイニング

相関関係を示した特徴ベクトルの格子データは、そのままではどの領域にどのような特徴がでているのか直感的に

理解しがたい。そこで特徴ベクトルを図 2 の (d)で示したように色で表現する。特徴ベクトルを色に割り当てるには、特徴ベクトルを 3 次元もしくは 4 次元に縮約し、適切な色空間の 1 点にマッピングする。色空間とは色を 3 次元もしくは 4 次元の数値ベクトルで表す方法であり、赤(Red)、緑(Green)、青(Blue)の RGB 手法や、色相(Hue)、彩度(Saturation)、明度(Brightness)の HSV 手法等がある。自動車のエンジン内の燃焼シミュレーションの例では、例えば温度が低く窒素濃度が高い、完全に燃焼が行われていない個所が特定の色で表示され、エンジン内のどこに発生しているかを知ることができると考えられる。このように 2 次元の格子状にデータを並べてその傾向を見る方法は Heat Map View と呼ばれており [4]、人間の目で直感的にどの領域でどのような特徴が出ているのかを大まかに知ることができ、調査する領域を絞り込むことが可能になる。

3. 本手法のバリエーション

本提案手法には、1)特徴ベクトルの定義、2)色空間へのマッピング方法の 2 点において、様々なバリエーションが考えられる。本節では各々の項目において考えられるバリエーションについて延べる。

3.1. 特徴ベクトルの定義

前節の例では 2 つの属性値の相関関係を特徴ベクトルにしていたが、同じシミュレーションにおけるパラメータを変化させたもの同士の相関関係を特徴ベクトルにする場合もある。いずれにしても特徴ベクトルとする空間の相関関係には大分して 1)対象となるデータセットやシミュレーションの目的に特化したもの、2)様々なデータセットに対して汎用的に適用できるもの、の 2 種類がある。1)は事例によって随時何が適切であるかが異なってくる特徴ベクトルである。例えば、前節の例では分布領域空間の空間関係が適切な特徴ベクトルだったが、流体シミュレーションの場合、流れ場を流線であらわしたときの形状や、その複数の流れ場が交わるなどの現象によって現れる特異点、物質が渦巻いていることを示す渦度、渦中心などそのシミュレーションを行っている専門家でないとは理解できないような特徴も数多く出てくる。これらの中から効果的な特徴ベクトルを決定するのは素人には難しい。それに対し 2)は、例えば位相、距離、方向関係など事例に関係なく適用できる特徴ベクトルになる。本稿ではこの 2) の様々な事例に対して汎用的に適用することを目的とし、2.1 節で説明したような空間的相関関係を用いた特徴ベクトルを生成する。具体的な特徴ベクトルの定義は 4 節にて述べる。

3.2. 色空間へのマッピング方法

ビジュアルデータマイニングを行う際、特徴ベクトルのどの要素をどの色空間の要素にマッピングするかを検討する必要がある。そもそも色空間には先ほど紹介した RGB モデルや HSV モデルのほかにも、シアン(cyan)、マゼンタ

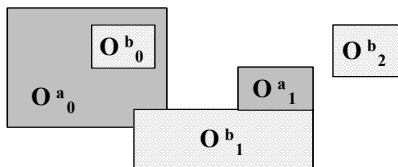
(magenta), 黒(black)の CMK, 輝度信号(Y)と二つの色差信号(CrとCb)の YCC など多数存在し, どれをどう使用すれば相関関係を見つけやすいのかは, 比較検討を繰り返して行く必要がある. ただし, 値が 360 度ある方向関係などには HSV 手法の色相(Hue)を当てはめるようにするなど, 特徴ベクトルの値の特性と色空間の要素の特性を近いものにする, マッピングをした際により特徴が分かり易いものが得られる.

4. 空間関係による特徴ベクトルの定義

3.1 節でも述べたように, 特徴ベクトルには事例によって異なるものと多くの事例に対して適用できるものがある. 本節では多くのシミュレーションデータに対して適用できる特徴ベクトルとして, 位相, 距離, 方向関係からなる特徴ベクトルの求め方について述べていく. 以降, 空間関係と特徴ベクトルの求め方を, 前述している自動車のエンジン内の燃焼の例を用いて説明する.

温度に関する条件 A によって得られる空間領域を S(A) と窒素濃度に関する条件 B によって得られる空間領域を S(B)とし, これらの空間関係 R(S(A), S(B))を特徴ベクトルとして求めることを考える.

データセットに対してある条件の空間領域を抽出した場合, 1 つのまとまった領域になるのではなく, 複数のまとまった領域に分散することが多い. そこでまず我々は空間領域 S(A)を複数の領域 {O^a₀, ..., O^a_n}で構成されているとする. 今回の例では, S(A)が 2 つのオブジェクト, S(B)が 3 つのオブジェクトを図 3 のような配置で有しているとする.



$$S(A) = \{O^a_0, O^a_1\} \quad S(B) = \{O^b_0, O^b_1, O^b_2\}$$

図 3 : S(A)と S(B)のオブジェクトの位置

次に, 複数の領域を持つ S(A_i) と S(B_j)との空間関係に関しては, 我々は文献[5]において提案されているシーン間の空間関係モデルに基づき空間関係 R(S(A), S(B))を, 位相関係 R_{Top}(S(A), S(B)), 距離関係 R_{Dist}(S(A), S(B)), 方向関係 R_{Dir}(S(A), S(B)) の三項で表す.

$$R(S(A), S(B)) = (R_{TOP}(S(A), S(B)), R_{DIR}(S(A), S(B)), R_{DIS}(S(A), S(B)))$$

以下, 各々の空間関係について述べる.

4.1.1. 位相関係

前項でも述べたように特徴ベクトルである S(A), S(B)間の空間関係を求めるためには S(A), S(B)間の位相関係を求める必要がある. S(A), S(B)はそれぞれ複数のオブジェクト

を有しているため, 実際の位相関係は S(A), S(B)のオブジェクトの全ての組み合わせについて位相関係を見ていくことになる. S(A)のオブジェクト O_a と S(B)のオブジェクト O_b 間の位相関係 R_{Top}(O_a, O_b)は OGC の 9-intersection モデルにより,

1. 離れている (Disjoint(O_a, O_b))
2. 外接している (Meet(O_a, O_b))
3. 一部分が重なっている (Overlap(O_a, O_b))
4. O_a が O_b に内接している (Covers(O_a, O_b))
5. O_b が O_a に内接している (CoveredBy(O_a, O_b))
6. O_a が O_b に含まれている (Contain(O_a, O_b))
7. O_b が O_a に含まれている (Inside(O_a, O_b))
8. 等しい (Equal(O_a, O_b))

の 8 種類に分類される. これらを利用することによって位相関係を二項関係として表現できる. 例えば図 3 の中で, オブジェクト O^b₀ は O^a₀ に含まれているので

$$R_{TOP}(O^a_0, O^b_0) = \text{Inside}(O^a_0, O^b_0)$$

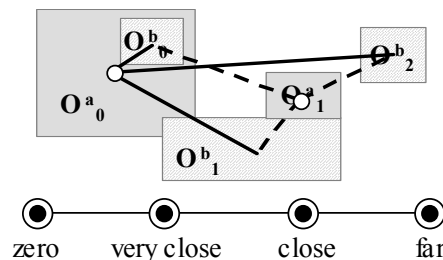
となる. 温度と窒素濃度の例 (図 3) の場合位相関係 R_{TOP}(S(A), S(B))は S(A), S(B)に含まれるオブジェクト間の全組合せによる 2 次元行列で表される (図 4).

	O ^b ₀	O ^b ₁	O ^b ₂
O ^a ₀	Inside	Overlap	Disjoint
O ^a ₁	Disjoint	Meet	Disjoint

図 4 : S(A)と S(B)の位相関係 R_{Top}(S(A), S(B))

4.1.2. 距離関係

S(A), S(B)間の距離関係も位相関係と同様, S(A), S(B)のオブジェクトの全ての組合せによる空間関係を求める. 2 オブジェクト間の距離関係 R_{Dist}(O_a, O_b)は O_a と O_b の実際の距離の値ではなく, 位相関係と同じように二項関係で表現するため, 実際の距離を大きく4段階に分けたものを使用する. O_a と O_b がとりうる最大の中心距離の値を近い方から very close, close, far に3分割し, 距離が0の場合は zero とする. 実際の距離がこれら4段階のなかのどれに当てはまるかを求め, 表現する. 温度と窒素濃度の例では, 距離関係を求めると図 5 の表のようになる.

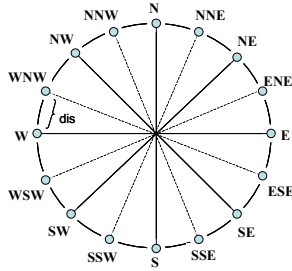


	O ^b ₀	O ^b ₁	O ^b ₂
O ^a ₀	Very close	close	far
O ^a ₁	close	Very close	close

図 5 : 距離関係 R_{Dis}(S(A), S(B))

4.1.3. 方向関係

方向関係についても全てのオブジェクトの組合せについて、オブジェクト間の方向関係 $R_{Dir}(O_a, O_b)$ を求める。また O_a, O_b の方向関係を実際の値で表すのではなく O_a を基準とした方向関係を二項関係で表現する。2次元の場合は図6の上図に表す18の方向関係に分類する。図3の方向関係 $R_{Dir}(S(A), S(B))$ を図6に示す。



	O^b_0	O^b_1	O^b_2
O^a_0	NE	ESE	E
O^a_1	WNW	SW	ENE

図6：方向関係とその距離 $R_{Dir}(S(A), S(B))$

今回の例では2次元のデータを用いているため、図6の方向関係は2次元になっているが、実際の実験で3次元データを用いる場合は図6の方向関係に高さの軸を加えて拡張したものを使用する。

4.2. 空間類似度を用いた色空間マッピング

4.1節にて定義した位相、距離、方向関係からなる空間関係 $R(S(A), S(B))$ を特徴ベクトルとし、これを色空間にマッピングすることにより、図2(d)に示されるような2次元図を生成する。本節では、特徴ベクトルを色空間にマッピングする方法について述べる。

前節の説明で分かるとおり $R(S(A), S(B))$ は二項関係の行列によって構成されている。一方色空間は3次元もしくは4次元の値によって表される。そのため、 $R(S(A), S(B))$ をそのまま色空間にマッピングするのは困難である。

そこで我々はある1組の条件 a, b による空間領域 A_0, B_0 の特徴ベクトル $R(S(A_0), S(B_0))$ を基準点とし、基準点との類似度を求めて、色空間に適用することにする。図7は $R(S(A_i), S(B_j))$ を色空間にマッピングするための手順を示している。基準点からの類似度は、位相関係、距離関係、方向関係において別々に求める。そして、それらの3種類の空間関係の類似度を0~1までの値になるように正規化した後、色空間の1点に割り当てる。図7の場合は、位相関係が10、距離関係が50、方向関係が129であった。これを正規化した結果 $D_{REG}(0.1, 0.5, 0.8)$ という3次元ベクトルが求められ、これをRGB空間の1点に割り当てて色を決定している。 D_{REG} とは空間類似度 D を正規化したものである。

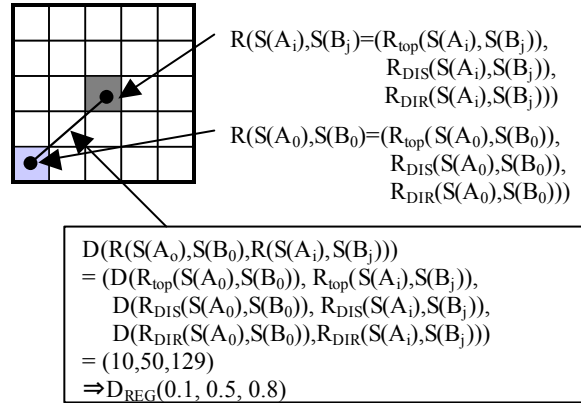


図7：空間類似度を用いた色空間マッピング

4.3. 空間類似度

我々は、 $R(S(A), S(B))$ と $R(S(A'), S(B'))$ との空間類似度 $D(R(S(A), S(B)), R(S(A'), S(B')))$ を以下のように定義する。

$$D_{REL}(R_{REL}(S(A), S(B)), R_{REL}(S(A'), S(B'))) = \sum D_{REL}(R_{REL}(O_a, O_b), R'_{REL}(O_a, O_b))$$

$$(O_a \in S(A) \cup S(A'), O_b \in S(B) \cup S(B'))$$

REL は空間関係の意味で TOP(位相関係)、DIS(距離関係)もしくは DIR(方向関係)のいずれかとなる。つまり、 $S(A)$ または $S(A')$ に含まれる任意のオブジェクト O_a と $S(B)$ または $S(B')$ に含まれる任意のオブジェクト O_b との空間関係 $R(O_a, O_b) (\in R(S(A), S(B)))$ と $R'(O_a, O_b) (\in R(S(A'), S(B')))$ との類似度の総和を空間類似度としている。

温度と窒素濃度の例では $R(S(A), S(B))$ について4.1節で求めてきたが、空間類似度を求めるためにもう1組み合わせの空間関係 $R(S(A'), S(B'))$ についても求めてあるとする。 $R(S(A'), S(B'))$ についての空間関係は図8の通りとする。

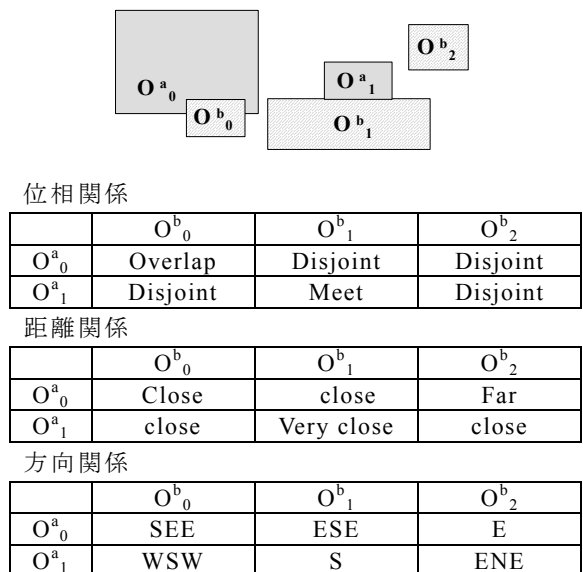


図8：空間関係 $R(S(A'), S(B'))$

ここで、 $R(S(A), S(B))$ を基準になる空間関係として $R(S(A'), S(B'))$ との類似度を求める。まず、同じオブジェクト間の関係において、二項関係が等しいものは類似度が 0 となるため削除し、異なる二項関係だけを残すと図 9 のように位相、距離、方向関係についての異なる関係が残る。次に二項関係同士の類似度を求める。これは文献[6]による空間関係の距離定義を下に算出する。位相の距離関係は図 10、距離の距離関係は図 5 の中図、方向の距離関係は図 6 の上図にそれぞれ記述されており、ある二項関係から別の二項関係にたどるための経路の数を距離とする。例えば位相 Covers と Inside の距離は equal を通る経路で、2 となる。同様に距離 VeryClose と far の距離は 2、方向 N と WSW の距離は 5 となる。

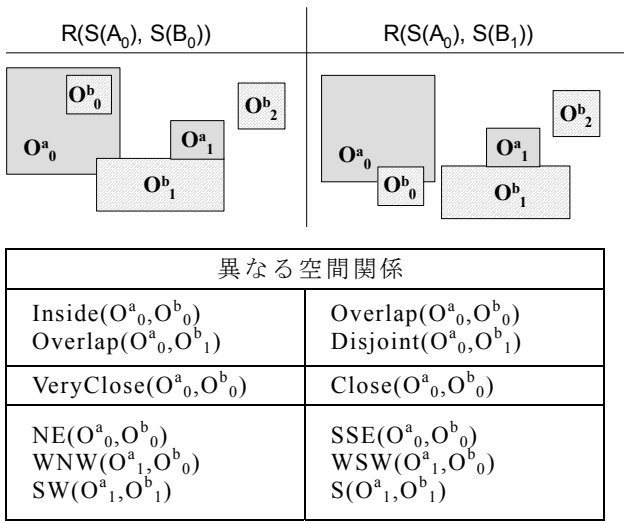


図 9：空間類似度の評価

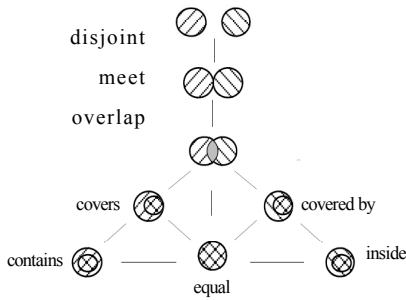


図 10：位相関係の距離関係

これらの距離を前述したようにそれぞれ測定していくと以下ようになる。

- $Dis(Inside, Overlap) = 2$
- $Dis(Overlap, Disjoint) = 2$
- $Dis(VeryClose, Close) = 1$
- $Dis(NE, SSE) = 6$
- $Dis(WNW, WSW) = 2$
- $Dis(SW, S) = 4$

これらの距離の総和をとることにより、位相関係の類似度 $D_{TOP}(R(S(A), S(B)), R(S(A'), S(B')))) = 4$ 、距離関係の類

似度 $D_{DIS}(R(S(A), S(B)), R(S(A'), S(B')))) = 1$ 、方向関係の類似度 $D_{DIR}(R(S(A), S(B)), R(S(A'), S(B')))) = 12$ を得ることができる。つまり全体の空間類似度は

$$D(R(S(A), S(B)), R(S(A'), S(B')))) = (4, 1, 12)$$

となる。

5. 提案モデルの妥当性検証

本稿で提案したモデルは、2 つのデータの相関関係における特徴をビジュアルマイニングによって発見するものである。そのためには特徴のないデータを本提案モデルに適用した場合は、各特徴ベクトルおよびカラーマッピング後のビジュアルデータにおいて何も特徴が出ない、つまり偏りなく値が分布する事が求められる。これは、もしここで特徴が出現してしまうと実際のデータに適用した際、出現した特徴がそのデータ固有の特徴なのか、あるいはノイズなのか判別できない可能性が出てくるためである。そこで我々は 4 節で提案した空間関係(位相、距離、方向関係)による特徴ベクトルとその類似度を求めるモデルを実際のデータに適用する前に、本提案モデルが実際に妥当であることを調べるためのシミュレーション実験を行った。

5.1. 手順

シミュレーションの大まかな手順は以下の通りである。

1. $1000 \times 1000 \times 1000$ の空間に、大きさと中心座標の異なる立方体のオブジェクトを 50 個ランダムに発生させたものを 2 つ生成する。
2. ランダムに発生させたオブジェクト同士の位相、距離、方向関係(各 50×50 個)を求める。
3. (2) で求めた空間関係の類似度を求める。
4. 1~3 を 100 回繰り返し類似度の値の分布を見る。

位相関係、距離関係の求め方については 4.1.1 節および 4.1.2 節において説明した通りである。またそれらの類似度についても 4.3 節で説明した通りに求めていく。ただし 4.1.3 節の方向関係の求め方は 2 次元空間に対しての定義だったため、今回新たに 3 次元空間に対する方向関係の定義を行う。3 次元の方向関係 $R_{DIR}(O_a, O_b)$ はオブジェクト O_a の位置、大きさを基に空間を 27 の領域に分類する(図 11) [7]。オブジェクト O_a が入っている領域は中心の領域(図 11 では 14)となり、オブジェクト O_a から見たオブジェクト O_b が 27 方向のうちどの領域に入るのかを求める。

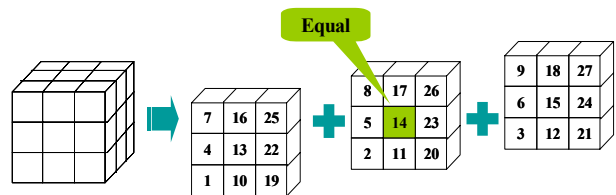


図 11：3次元における方向関係の方向分割

空間類似度を求めるため各方向の距離関係は隣り合う領域同士の距離を 1 とする。例えば方向 1 (dir1) と方向

23 (dir23) の距離 $\text{Dis}(\text{dir1}, \text{dir23})$ は、1 から 23 に行くための最短経路の距離が {1-10-19-20-23} というように 4 となるため、 $\text{Dis}(\text{dir1}, \text{dir23})=4$ である。

5.2. 結果と考察

空間関係(位相, 距離, 方向)の類似度を求める計算を 100 回試行した結果の分布を図 12~13 に示す。

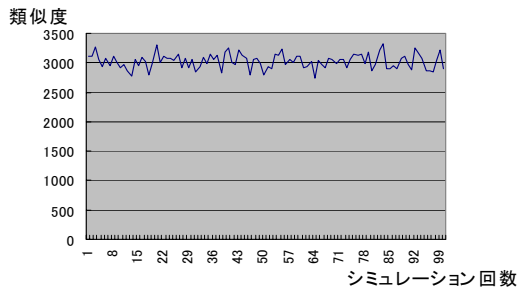


図 12 : 位相関係の類似度

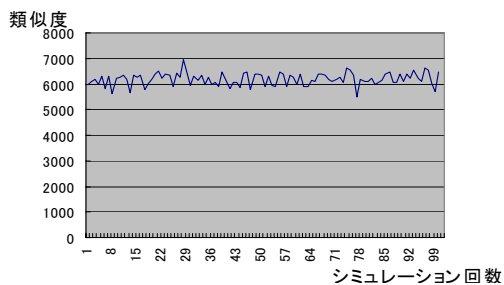


図 13 : 距離関係の類似度

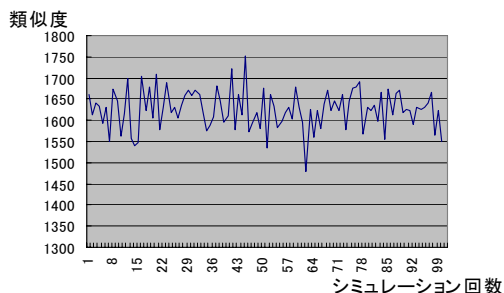


図 14 : 方向関係の類似度

図 12~13 から明らかなように位相, 距離, 方向関係の類似度はいずれも特徴となるような値が出ていない。よって本提案モデルを特徴のないデータに適用した場合は, 確かに特徴が出ないことが分かった。

6. 実験

本節では特徴が判断し易いテストデータを作成し, 4 節で述べた提案手法を適用した際特徴が正しく可視化されているかどうかを確認する。

6.1. テストデータ

今回特徴を判断し易いデータとして 2 つのテストデータ (data1, data2) を作成した。data1 は $x = 50, y = 50, z = 50$ の格子状空間の $(x, y, z) = (0, 0, 0)$ から $(x, y, z) = (50, 50, 50)$

に引いた対角線にそって球をランダムに発生させたものである。ランダムに発生した球は 1 以上 50 未満の値を持っており, その値は $(x, y, z) = (0, 0, 0)$ に近いほど大きく, $(x, y, z) = (50, 50, 50)$ に近づくにつれて徐々に小さくなっていく。data2 も同様の格子状空間で $(x, y, z) = (50, 0, 0)$ から $(x, y, z) = (0, 50, 50)$ に向かって値が小さくなっていく物体をイメージして作成したデータである(図 15)。温度は両データとも最大値が 49, 最小値が 0 で, 2 つのデータは格子状空間の中心付近で交わっている。

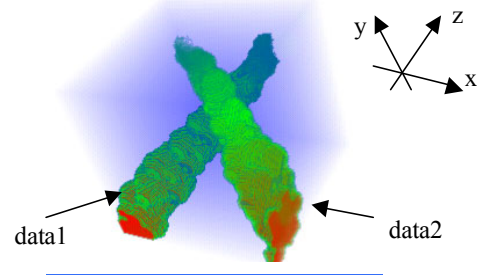


図 15 : 作成したテストデータ (data1, data2)

6.2. 実装

実装手順として, 始めにある一定の範囲の値とその座標を効率よく取得するために, 作成したデータ data1, data2 をデータベースに入れる。今回カラーマッピングを行うのは data1, data2 を軸とする 50×50 の 2 次元格子空間である。それらの格子 1 つ 1 つに入る値(類似度)を求めるためまずは空間関係を求める。空間関係を求めるために必要な各シーン内のオブジェクトは, あるシーンのデータをデータベースから取得した後クラスタリングを行う事によって定義できる。クラスタリングをする事によってデータはまとまった領域ごとに分割され, その領域を全て含むような直方体 MBB (Maximum Bounding Box) をオブジェクトとする。今回の実装では data1 の各シーンを $S(A)$ とすると $S(A_0)$ は 0 以上 1 未満のデータの集合, $S(A_1)$ は 1 以上 2 未満のデータの集合になる。同様に data2 の各シーンを $S(B)$ とすると $S(B_0)$ は 0 以上 1 未満, $S(B_1)$ は 1 以上 2 未満のデータの集合になる。例えば $S(A_{45})$ と $S(B_{50})$ のオブジェクトを求める場合は, data1 における $45 \leq \text{値} < 46$ の等値面と data2 における $48 \leq \text{値} < 49$ の等値面が図 16 のようになるのでオブジェクトはそれらを覆う直方体 (MBB) になる(図 16 の O^a_0, O^b_0)。

上記のようにシーン内のオブジェクト定義した後, 2 シーン間の空間関係 $R(S(A), S(B))$ を求め, さらに $R(S(A_0), S(B_0))$ とその他の空間関係 $R(S(A_n), S(B_m))$ 間の類似度 $D(R(S(A_0), S(B_0)), R(S(A_n), S(B_m)))$ を求めていく。そして位相関係の類似度 $D_{\text{TOP}}(R(S(A_0), S(B_0)), R(S(A_n), S(B_m)))$ を R , 距離関係の類似度 $D_{\text{DIS}}(R(S(A_0), S(B_0)), R(S(A_n), S(B_m)))$ を G , 方向関係の類似度 $D_{\text{DIR}}(R(S(A_0), S(B_0)), R(S(A_n), S(B_m)))$ を B とする RGB 値 ($0 \leq R, G, B \leq 255$) をマイニング用の格子状空間にカ

ラーマッピングしていく。

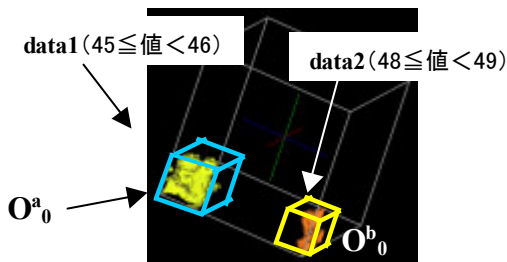


図 16 : data1(45 ≤ 値 < 46), data2(48 ≤ 値 < 49) の等値面

6.3. 結果と考察

テストデータ data1, data2 に本提案手法を適用した結果, 図 17 のようなビジュアルマイニング用の可視化データが得られた。

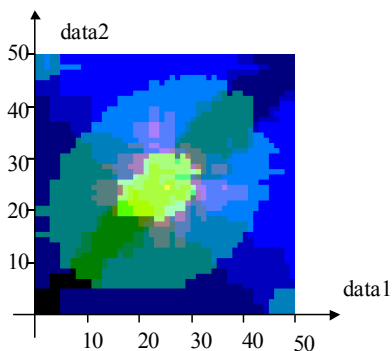


図 17 : 本提案手法適用結果

カラーマッピングの値である類似度は全て, 基本となる空間関係 $R(S(A_0), S(B_0))$ とその他の空間関係 $R(S(A_n), S(B_m))$ との類似度であり, 2 つの空間関係が類似している場合は黒に近く, 逆に類似していない場合は明るい色になる. 図 17 の結果では, data1, data2 の値が 20~30 の間で色が極端に明るくなっているところから, この値において何かしら相関関係に特徴があるのではないかと推測できる. そこで実際のデータの等値面画像(図 18)を見てみると, この 20~30 の値において 2 つのデータが交差していることが分かる。

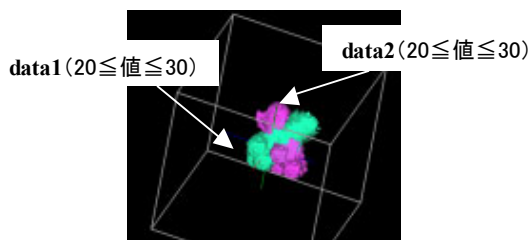


図 18 : data1(20 ≤ 値 ≤ 30), data2(20 ≤ 値 ≤ 30) の等値面

さらに方向関係を表す青色が $y = -x + 50$ の線を境に明るさが反転している事から, オブジェクトの位置関係が逆になっ

ている事が推測できる. これも図 15 を見ると実際にオブジェクトの方向関係が逆になっていることが分かる。

以上の事からテストデータ data1, data2 の相関関係の主だった特徴がどの範囲の値に出ているかは, 図 17 の可視化データからビジュアルマイニングによって大まかに推測する事ができることが分かった。

7. まとめと今後の課題

本稿では複数の要素の分布における空間的關係や, 複数のパラメタ間による物理現象の相関關係を発見する手がかりを得るためのマイニング手法を提案した. 本提案は空間的相関關係を表現する特徴ベクトルを生成し, それを格子状空間に並べカラーマッピングする事によって, 空間的相関關係のオーバービューを作成しビジュアルマイニングを行うものである. 特徴ベクトルは対象データや目的によって大きく異なってくるため, 我々は今回比較的汎用性のある空間關係を特徴ベクトルとするものを提案した. そしてその提案モデルが有用であるかどうかを調べるためにシミュレーション実験を行った. さらに特徴の出やすいデータを本提案手法に適用し, 確かに特徴が出る事を確認した. これにより, 人間の目で直感的にどの領域でどのような特徴が出ているのかを大まかに知ることができ, 調査する領域を絞り込むことが可能になると考えられる。

今後の課題としては, さらに複雑な特徴をもったテストデータを作成し, 本提案手法を適用する事によって正しく特徴がでるかを確認する. その後実際のシミュレーションデータに本手法を適用して, さらに汎用的に使用できるよう本手法を改良していく予定である。

文 献

- [1] 藤原和人, 松尾日出男, 小野謙二, 広江哲幸, "くいちがい配置された2正方形柱まわりの流れ", 日本航空宇宙学会誌, Vol.40, No.460, pp.276-284, 1992.
- [2] OCG モデル : <http://www.opengeospatial.org/>
- [3] M. Egenhofer and R. Golledge "Spatial and Temporal Reasoning in Geographic Information Systems", Oxford University Press, New York, p.304, 1998.
- [4] Usama Fayyad, Georges G. Grinstein and Andreas Wierse (editors), "Information visualization in Data Mining and Knowledge discovery", Morgan Kaufmann Publishers, p.407, 2001.
- [5] Tom Bruns and Max Egenhofer, "Similarity of Spatial Scenes", Seventh International Symposium on Spatial Data Handling (SDH '96), Delft, The Netherlands, M.-J. Kraak and M. Molenaar (eds.), pp. 4A.31-42, August 1996.
- [6] M. Egenhofer and D. Mark, "Modeling Conceptual Neighborhoods of Topological Line-Region Relations", International Journal of Geographical Information Systems 9 (5): 555-565, 1995.
- [7] Max J.Egenhofer, "Spetial Relations: Models, Inferences, and their Future Applications", Proceedings of Advanced Database Symposium '96, p.9, 1996, December.