時空間変動データからの ホットスポット自動抽出・要約システムの開発

本田 理恵 [*1], 林 諒 [*1]

Development of Automatic Extraction and Summarization System for Hot spots in Spatio-Temporal Data

Rie HONDA[*1] and Ryo HAYASHI[*1]

Abstract

"The hot spots" which have the values unlike ones in the neighborhood often appear in the spatio-temporal data set in the field of space and planetary sciences. Method of automatic extraction of such hotspots as objects from the spatio-temporal data by using mixture distribution of multivariate normal distributions are described, which aims to understand changing patterns such as co-occurrence or temporal rules among the hot spots and to predict them. This paper also introduces the example system in which hotspots are extracted from meteorological satellite imagery. Plans for future work are also discussed based on the problem that became clear through this study example.

Keywords: spatio-temporal data, hot spot, object extraction, pattern discovery, mixture distribution

概要

宇宙，惑星科学の分野で取得される時空間データには，周囲とは異なる値を持つ“ホットスポット”に着目すべき問題が現れることが多い．このホットスポット領域をオブジェクトとして自動的に抽出して記載し，その時空間変動パターンを抽出することができれば様々な問題に応用出来ると考えられる．本研究では時系列グリッドデータから多変量正規分布の混合分布でモデル化し，そのモデルパラメータを求めることによってホットスポットを自動抽出する手法と，これを使って気象衛星画像からインタラクティブ知識発見システムの構築例について紹介し，これらの研究例を通して明らかになった問題から今後の方向性について述べる．

1. はじめに

近年諸分野で大量の時空間データが蓄積されるようになっている．日常生活ではセキュリティカメラの画像などが典型的な例であるが，地球観測衛星や宇宙惑星科学分野でもリアルタイムモニタリングされた様々なデータが蓄積され，時間とともに成長する巨大な時空間データのアーカイブを形成している．こうした時空間のビッグデータから変動パターンを抽出し，自然・社会現象の変動パターンのモデルを推定することができれば，現象の理解や予測に利用することができる．

このような時空間データには周囲と異なる値をもついわゆる”ホット(あるいはコールド)スポット”が含まれることが多い．気象衛星画像であれば雲の存在領域が，太陽観測衛星のデータにおいては太陽黒点などがこれにあたる．他にも地球観測衛星で得られる植生指標の異常値領域，地震学的なトモグラフィーで得られる地球内部の高温・低温領域もこれにあたる．流体力学のシミュレーションによって生成された様々な物理量（温度，渦度など）もホットスポットの典型例を含む．

このような時空間データは，空間的には２次元，すなわち時系列画像であることが多かったが，最近では３次元データの空間データも取得されるようになっている．３次元シミュレーションの物理量の記録はこの典型例であるが, 観測の分野でもフェーズドアレイ気象レーダによってリアルタイムの３次元の反射強度(降水コアの存在を示す）が取得されるようになっている[1].

本論文では，このホットスポットに基づいて時空間データの要約と知識発見を行う問題について，気象衛星画像やフェーズドアレイ気象レーダデータに対して，筆者らが検討しているオブジェクト抽出システムの検討内容を汎用的な時空間データからのホットスポット抽出の観点からまとめて報告し，そこに見られる問題点を整理して，今後の開発の方向性を探ることを目的とする．

doi: 10.20637/JAXA-RR-17-009/0004
[*] 平成 30 年 1 月 18 日受付（Received January 18, 2018）
[*1] 高知大学（Kochi University）

2. 手法

2.1 ターゲット，タスク定義

最初にターゲットとなるデータの性質について定義する．ターゲットは２次元，または３次元のグリッド状に配置されたフィールド値の時系列データとする．このフィールド値は気象画像では輝度，フェーズドアレイ気象レーダでは反射強度などの物理量を反映した値にあたる．実世界の様々な分野では，まばらにおかれたセンサや移動するセンサから時間とともに取得されるデータなど非グリッド状のデータも普遍的に存在するが，ここでは簡単化と，実際に地球観測衛星や宇宙惑星科学データで取得されるデータは最終的にはグリッドデータの形に変換されることが多いという理由から，対象をグリッドデータに限るものとする．

2.2 ホットスポットの定義

抽出対象であるホットスポットは以下のような特徴をもつものとする．

(1) 周囲と異なる値をもつ塊状の領域として現れる．
(2) 誕生から消滅に至る生存期間をもつ．
(3) ２つ以上のホットスポットは重なって存在しうる．
(4) ホットスポット同士は分裂，融合などの相互作用を行い得る．

これは気象画像での台風等の自然界で発生するホットスポットを想定したものとなっている．

こうしたホットスポットは“オブジェクト“の１種として捉えることができる．”オブジェクト”という言葉は画像認識では物体を表すものとして，またオブジェクト指向では属性と手段をもつ実体を意味するものとして用いられる．本研究ではホットスポットをオブジェクトとして表現することで，このような特徴を持つものとしてホットスポットを特徴付ける．例えばホットスポットの抽出後にはその細かい特徴（サイズ，分散，あるいはパターン内部のテクスチャ）はオブジェクトの属性として記載する．以降では一般論としてのオブジェクトとホットスポットを必要に応じて使い分ける．

2.3 多変量正規分布の混合モデルによるオブジェクト抽出手法（標準法）

まず図１のように塊状に分布する M次元のデータ集合$\boldsymbol{D}$の塊をオブジェクトとして表現する事を考える．

$$\boldsymbol{D} = \{\boldsymbol{d}_i \in R^M \mid i = 1,2,3,\cdots,n\} \tag{1}$$

ここでは重なって存在する不特定の形状を扱うため，オブジェクトを多変量正規分布で表現し，この混合分布で全体を表現するものとする．観測事例の分布を多変量正規分布の混合モデルで近似する手法は機械学習分野でのクラスタリングに一般的に用いられる手法[2]でもある．この手法について以下に説明する．

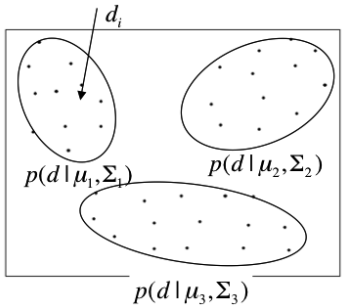


図 1. オブジェクトの多変量正規分布の混合分布近似（M=2 の場合）

図１の１点$\boldsymbol{d}$を観測する確率密度分布が下記の通り 1 つのオブジェクトの存在を表す多変量正規分布でモデル化できるものとする．

$$p\left(\boldsymbol{d}\middle|\boldsymbol{\mu}, \Sigma\right) = \frac{1}{\left(\sqrt{2\pi}\right)^M \sqrt{|\boldsymbol{\Sigma}|}} \exp\left\{-\frac{1}{2}(\boldsymbol{d}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{d}-\boldsymbol{\mu})\right\} \tag{2}$$

ここで$\boldsymbol{\mu} \in \boldsymbol{R}^M$は多変量正規分布の中心ベクトル（平均値），$\boldsymbol{\Sigma}$は$M \times M$の分散共分散行列とする．$\boldsymbol{\mu}$はオブジェクトの中心，$\boldsymbol{\Sigma}$はオブジェクトの広がりを表す．複数のオブジェクトが存在しうる時，座標$\boldsymbol{d}$においてオブジェクトを観測する確率密度分布は複数成分からなる多変量正規分布の重み付きの平均で表すことができる．

$$P(\boldsymbol{d}) = \sum_{j=1}^{K} \omega_j \, p\left(\boldsymbol{d} \middle| \, \boldsymbol{\mu_j}, \boldsymbol{\Sigma_j}\right) \tag{3}$$

ここでω_jは重み係数$(\sum_{j=1}^{K} \omega_j = 1)$，$K$は成分数を表す．このパラメータをまとめて以下の通りとする．

$$\boldsymbol{\theta} = \{(\boldsymbol{\mu_j}, \boldsymbol{\Sigma_j}, \omega_j) | \, j = 1, \ldots . K\} \tag{4}$$

観測値$\boldsymbol{D}$は，それぞれどの多変量正規分布から発生したものかを知ることができないいわゆる不完全データであり，最尤推定などの手法で直接パラメータを求めることができない．このため，EM（Expectation and Maximization ）アルゴリズム[3]でパラメータを求める．EM アルゴリズムは，仮に与えたパラメータ$\boldsymbol{\theta}$からモデルを計算し，さらにモデルの対数尤度を増加させるような新しいパラメータ$\boldsymbol{\theta'}$を推定するという操作を繰り返すことによって対数尤度$L(\boldsymbol{\theta})$を最大化するパラメータ$\boldsymbol{\theta'}$を求めるものである．

$$L(\boldsymbol{\theta}) = \log\left(\prod_{i=1}^{n} P\left(d_i | \boldsymbol{\theta}\right)\right) \tag{5}$$

$$\boldsymbol{\theta'} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \, L(\boldsymbol{\theta}) \tag{6}$$

まず標準的な方式での EM アルゴリズムによるパラメータの計算手順[4]を示す．

<EM アルゴリズムによるパラメータ推定>
E ステップで寄与率を推定し，これをもとに M ステップで対数尤度を最大化するパラメータを推定し，この値を用いて寄与率を推定するというプロセスを収束するまで続ける．ここで寄与率$z(i,j)$は，観測値$\boldsymbol{d}_i$がj番目の多変量正規分布（オブジェクト）から発生したことを示す推定量である．

1. $\boldsymbol{\theta} = \{(\boldsymbol{\mu}_j, \Sigma_j, \omega_j) | j = 1,2, \cdots, K\}$に初期値を与える．
2. E-Step：寄与率$z(i,j)$を計算する．このとき変化 $z(i,j) - z'(i,j)$ が小さければ終了．そうでない場合 3 を実行する．

$$z'(i,j) = \frac{\omega_j p(\boldsymbol{d}_i | \, \boldsymbol{\mu_j}, \boldsymbol{\Sigma_j})}{\sum_{k=1}^{K} \omega_k p(\boldsymbol{d}_i | \, \boldsymbol{\mu_k}, \boldsymbol{\Sigma_k})} \tag{7}$$

3. M-Step：対数尤度を最大化させるようモデルパラメータを更新し，$z(i,j) \leftarrow z'(i,j)$として2 に戻る．

$$\omega_j{}' = \frac{\sum_{i=1}^{n} z'(i,j)}{n} \tag{8}$$

$$\boldsymbol{\mu}_j{}' = \frac{\sum_{i=1}^{n} z'(i,j) \, \boldsymbol{d}_i}{\sum_{i=1}^{n} z'(i,j)} \tag{9}$$

$$\boldsymbol{\Sigma}_j{}' = \frac{\sum_{i=1}^{n} z'(i,j) \left(\boldsymbol{d}_i - \boldsymbol{\mu}_j\right)\left(\boldsymbol{d_i} - \boldsymbol{\mu_j}\right)^T}{\sum_{i=1}^{n} z'(i,j)} \tag{10}$$

なお成分数Kについては総当たりで上記の手法でもとめたモデルパラメータに対してベイズの情報量基準(BIC)[5]を用いて，最適な(すなわちこれを最小化する)Kを求める．Fは自由度である．

$$\mathrm{BIC} = -2nL(\boldsymbol{\theta}) + F\log n \tag{11}$$

2.4 フィールド値をもつグリッドデータからのホットスポット抽出への応用

2.3 の手法をグリッドデータに適用するためには，単純には観測値が閾値をこえた座標をサンプリングするアプローチをとることができる[6][7][8]．M次元のグリッドの座標$\boldsymbol{d}_i$に存在するフィールドデータ$I(d_i)$に対してある閾値I_{th}を超えた座標$\boldsymbol{d}_i$をホットスポットの存在する点の座標集合$\boldsymbol{D}$としてサンプリングして使用する．

$$\boldsymbol{D} = \left\{ \boldsymbol{d}_i \middle| I(\boldsymbol{d}_i) > I_{th}, i = 1,2,3,\cdots,n \right\} \tag{12}$$

ただしこの手法では 1 座標(グリッド点)に対して観測点が 1 点となってしまい，生成されたデータは平坦な確率密度分布を示す事になり，モデルとした多変量正規分布からの乖離が大きくなってしまう．気象レーダの反射強度などでは塊の中心にむかってその値が増加し，フィールド値自身が多変量正規分布に類似した分布をしている場合も多いため，問題によっては観測値そのものを有効活用したほうが良い．

以下では気象レーダデータのようにフィールド値自身が多変量正規分布に調和的であり，中心ほどその値が高いというケースを考える．この場合，観測値の大きさ自体が観測頻度に比例すると仮定することによって，フィールド値の分布により適合したモデルパラメータを求めることができる．あらかじめバックグラウンドノイズやオフセットの影響を取り除くため，フィールド値に対して，閾値I_{th}を差し引くオフセット処理を行う．

$$\delta I_i = \max\left\{ I(\boldsymbol{d}_i) - I_{th} \right\} \tag{13}$$

この項を利用して，2.3 の標準法(8)，(9)，(10)式の$\sum_{i=1}^{n} 1$を$\sum_{i=1}^{n} \delta I_i$で置き換えることによって，フィールド値が確率密度を反映するという効果を簡易に取り込むことができる．以下にその変更部分のみを示す．

<EM アルゴリズムによるパラメータ推定．フィールド値を活用する場合の 3. M-Step 変更部分のみ抜粋>

$$\omega_j{}' = \frac{\sum_{i=1}^{n} z'(i,j)\,\delta I_i}{\sum_{i=1}^{n} \delta I_i} \tag{14}$$

$$\boldsymbol{\mu}_j{}' = \frac{\sum_{i=1}^{n} z'(i,j)\,\boldsymbol{d}_i \delta I_i}{\sum_{i=1}^{n} z'(i,j) \delta I_i} \tag{15}$$

$$\boldsymbol{\Sigma}_j{}' = \frac{\sum_{i=1}^{n} z'(i,j)\left(\boldsymbol{d}_i - \boldsymbol{\mu}_j\right)\left(\boldsymbol{d}_i - \boldsymbol{\mu}_j\right)^T \delta I_i}{\sum_{i=1}^{n} z'(i,j)\,\delta I_i} \tag{16}$$

多変量正規分布の混合分布から生成した 3 次元の模擬データを用いた実験[9]では，この変更を施すことによって，標準手法では縮退や分裂を起こして正しい成分数を求める事ができなかったケースでも，正しい成分数の解が求められるようになり，モデルとデータの一致度が大きく改善したことを確認している．

なお，この手法はそれぞれデータの性質によって選択して使用されるべきものである．2.6 で述べる気象画像からの雲塊の抽出などの問題では，中心ほど信号が高いわけではなく，閾値処理によって抽出されたサンプリング点の分布自体がモデルパラメータの決定において重要である．このような場合は従来のフィールド値を使わない標準手法が適当である．一方，気象レーダの反射強度など中心付近にむかって信号が大きくなる性質のある問題においては，フィールド値が確率密度分布モデルの値に比例することを仮定する今回の手法(2.4 節)を用いる事が適当であると考えられる．

2.5 ホットスポットの消滅，融合，生成

ここまで述べた手法で，ある時間断面で抽出されたホットスポットは，一定の期間存続し，やがて消滅する．また生存期間のうちに他のホットスポットと融合したり，分裂してあらたなホットスポットを生成したりする．この過程を表現するために，パラメータの抽出過程において前の時間の解を次の時間に継承させる手法を提案している[7][8]．具体的には前の時間の解をラベルとともに次の時間の解の初期値に継承させるという単純な手続きによって，時間的な連続性をもってホットスポットの追跡を行うことを可能としている．一方，消滅や分裂が発生する可能性も考慮して，解の継承の際に下記のような複数の初期値を与えて，並列（下記の場合 *2q+1* ケース）にモデリングし，BIC に従って最良の解を選択する手法をとっている．

(1) 前の時間の解そのものを初期値として与える(成分数 *K*，1 ケース)
(2) 前の時間の解から重み係数の低い順に解を *q* 個まで消去して初期値として与える(成分数 *K-1, …, K-q*, *q* ケース)
(3) 前の時間の解から重み係数の大きい順に解を *q* 個までそれぞれ 2 つに多重化して初期値として与える(成分数 *K+1, …, K+q*, *q* ケース)

実際に気象画像や気象レーダデータに適用した例では，時間ステップで分布が大きく様変わりすることなければ，*q* は 2-3 で適当なことが多い．また，ここでは新たに別の箇所にホットスポットが発生するケースは陽には扱っていないが，実際には(3)の過程で分裂用に生成された初期解が，この新規発生の成分へ移動して

発見されていることが多い(2.6，図 2 参照)が，この弊害として”新規発生”と”分裂”が混同されて扱われてしまうことになる．このような影響についても，事前，事後に判定するプロセスを付加する等，今後より詳細な検討を進めていく予定である．

2.6 ホットスポット抽出の実例

本章で述べたホットスポット抽出手法の実データへの適用例として，気象衛星ひまわり IR 画像（MTSAT-7，高知大気象情報頁[9]）からの雲塊抽出の例を紹介する．ターゲットデータは２次元画像であり，画像の輝度が高い部分に雲が存在する．これがホットスポットであるとみなして，この手法を適用する．この場合は雲塊の中心ほど輝度が高いということはないため，閾値によるサンプリングのみ用いる標準的な手法(2.3 参照)を使用する．

図２に 6 時間おきの画像に対して，求められたモデルの多変量正規分布のカラーコンターとそのラベルで抽出結果を示す．図 2 下の得られたモデルの各成分に付与されている ID は 2.5 でのべた手順でつけられた同一性を示すラベルである．このうち図 2 下中央の 7-1，7-2 のような番号はその前の時間で１つであった 7 番のホットスポットから分裂したことを示している．全般的には，発見的に雲の塊を個数も含めて決定できていることがわかる．また，時間を追って同一のホットスポットを追跡できていることもわかる．一方で最初の時間から次の時間に移行する過程で７番目の成分が２つに分裂したとみなされているが，この１つは実は新規発生であって分裂を示している訳ではない．これは 2.5 で述べた通り現在のフレーワークには新規発生が含まれていないためであり，今後の検討が必要である．

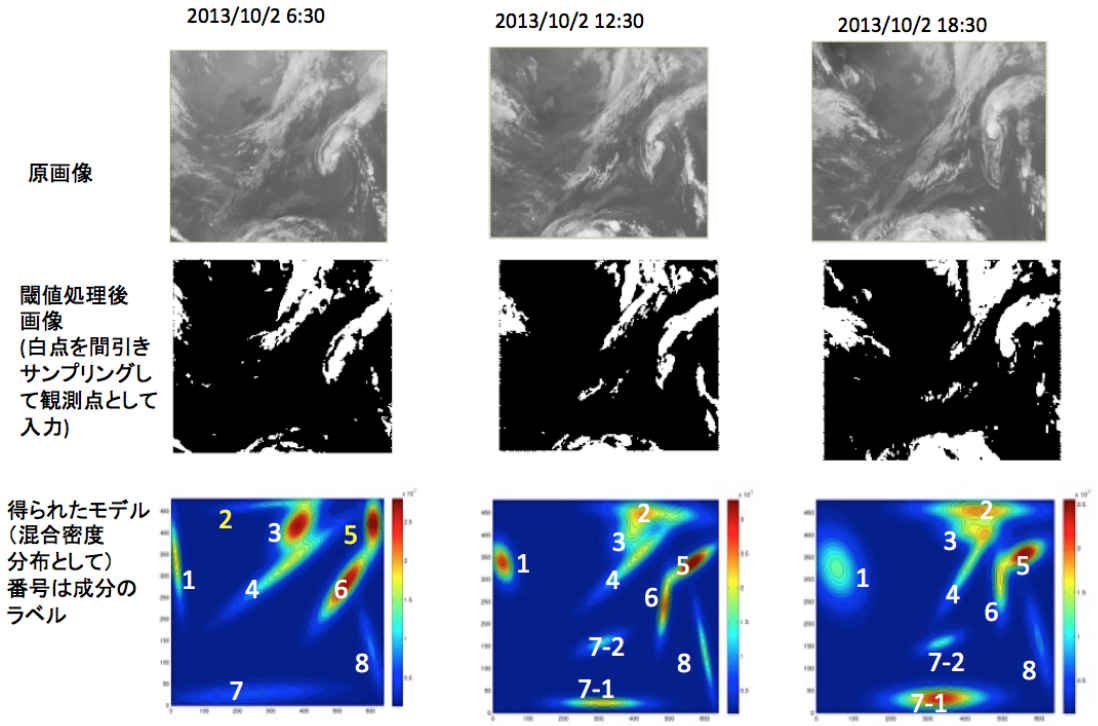


図 2. 気象衛星画像(2013 年 10 月 2 日)からのホットスポット抽出の例.

次に 3 次元データへの適用例として，フェーズドアレイ気象レーダで観測した反射強度データからの降水コア抽出[1]への適用例を示す．図 3 の例では赤みがかった部分ほどレーダの反射強度が大きく，この塊状の領域が濃い雨雲のある降水コアにあたる．降水コアの中心に向かって反射強度の値が大きくなっていることと，この降水コアの上昇や下降が豪雨の発展過程と関わっていると考えられることから，2-4 の強度の重みを使用する手法によるモデリングの精度の向上が有効と考えられる．

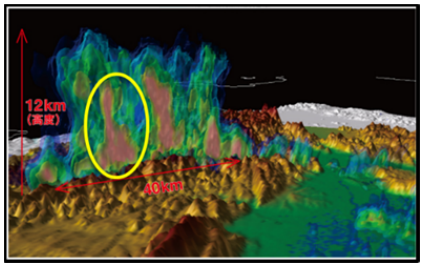


図 3 フェーズドアレイ気象レーダデータに現れた雨雲とホットスポットの例[1]

図 4 に実際に 2012 年 7 月 16 日大阪大学吹田キャンパス付近で発生した豪雨をフェーズドアレイ気象レーダで観測した反射強度データに対して本手法を適用した例を示す[10]．ここでは閾値を 30dBZ として座標のみを取り出して 2.3 の標準的な手法を使用したケース（図 4 左）と，閾値処理後フィールド値（反射強度）を重みとして 2.4 の改良手法を使用したケース（図 4 右）の比較を示す．簡単のため各観測点において寄与率が最大の成分毎に同じ色で示すことで，抽出された成分の分布を示している．座標のみを取り出して入力値をしたケース（図 4 左）では，特に右上の大きな塊が過剰に細かい成分に分裂しているが，フィールド値を用いた手法（図 4 右）では，この分裂が抑制され，自然な分布に近づいていることが観察できる．よって予備的には中心に向かってフィールド値が大きくなるようなグリッドデータからホットスポットをオブジェクトとして抽出する問題においては，フィールド値を使用することが効果的であることが示される．

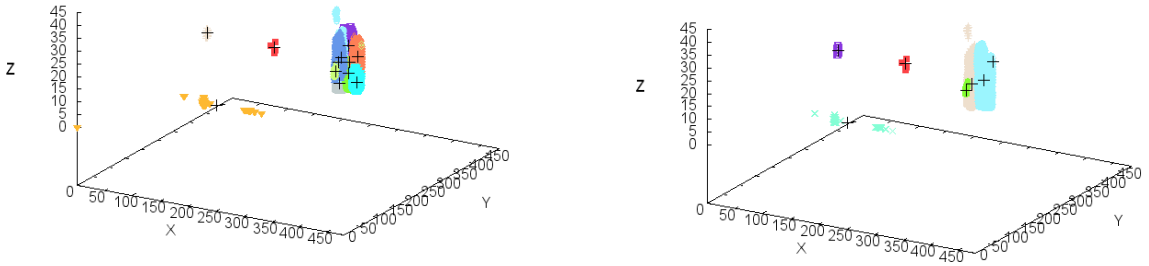


図 4.　2012 年 7 月 16 日 17 時 00 分の大阪大学吹田キャンパスのフェーズドアレイ気象レーダ[1]で観測した反射強度のモデリングの結果得られた多変量正規分布の成分の分布．左は敷居値処理(閾値 30dBZ)によって座標のみを入力値として使用した結果で，右は敷居値処理後，強度による重み付けを行った結果[10]．各点を寄与率が最大の成分毎に同じ色で表現している．また黒い十字は各成分の重心である．

3　時空間ホットスポット抽出，要約システム

3.1 システムの概要

ここまでに述べたホットスポットの抽出手法をコアプロセスとして，汎用的な時空間ホットスポット抽出，要約システムの設計について述べる．満たすべき要件は以下の 4 つになる．

(1) ホットスポット＝オブジェクトを自動的に時系列２次元，３次元グリッドデータから抽出できること
(2) ホットスポットの特徴に応じて意味的ラベル付けができること
(3) これらの情報がデータベースに格納され，簡易に検索できること
(4) これらの情報を活用して，ホットスポットの相互作用や変遷を時間，空間の両面で要約，可視化できること

抽出されたホットスポットには２で述べられた手法を適用することにより，中心位置，広がり，存続期間，親子関係のラベルなどが自然に付帯するが，そのホットスポットが何を表しているのかという意味的なラベル(2)がさらに付与されると，次の段階での知識発見における有用性がさらに増す．この意味的ラベル付けにはパターン認識，機械学習のアルゴリズムを用いる．先行研究[7][8]では Kohonen の自己組織化マップ[11]による教師なし学習によるクラスタリングに基づくラベル付けを行っているが，教師データを積極的に利用して，Deep Learning などのより精度の高い分類アルゴリズムを用いてもよい．

これを実現するシステムのイメージ図は図 5 のようになる．意味付けの箇所は必要に応じて様々なアルゴリズムで置き換えることが想定される．またユーザとのインターフェースは実際には web sever を通じて行うことを想定する．

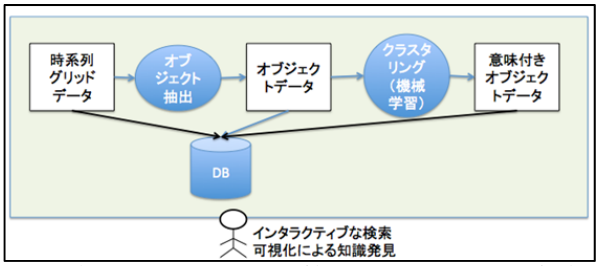


図 5　時空間ホットスポット抽出，要約システムの概念図

3.2 データモデル

このようなシステムを構築するにあたっては オブジェクトについてのデータモデルを整理しておく必要がある．図 6 に松永ほか(2016)[8]の検討結果を示す．ここに示されているように，ここまで抽出対象としてきた

オブジェクトは”スナップショット的な画像内のみかけのオブジェクト”であり，その背後で時間的に変化しながら存続し続ける”真のオブジェクト”の時間断面にすぎない．またオブジェクト間には分裂によって生じる親子関係としてファミリーに相当するものが存在する．実際にはこの３つは図 6 下の図の概念図のように階層的な構造で表現可能である．なおこの概念図ではオブジェクトの融合についてはまだ含まれていない．

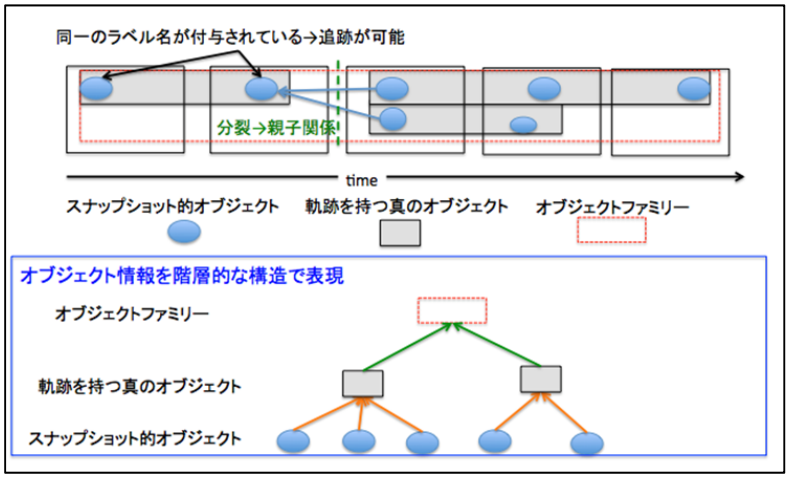


図 6 オブジェクトの階層構造([8]に基づく)

図 7 にはさらにこの３つの階層のオブジェクトのモデルを UML (Unified Modeling Language) のクラス図で示す．”真のオブジェクト”は”スナップショットオブジェクト”の集約として，また”オブジェクトファミリー”はこの真のオブジェクトの集約として定義される．真のオブジェクト，オブジェクトファミリーの情報は，２で述べたオブジェクト抽出手法で自然に得られるスナップショット的オブジェクトの情報から再構成され，データベースに格納されるものとする．

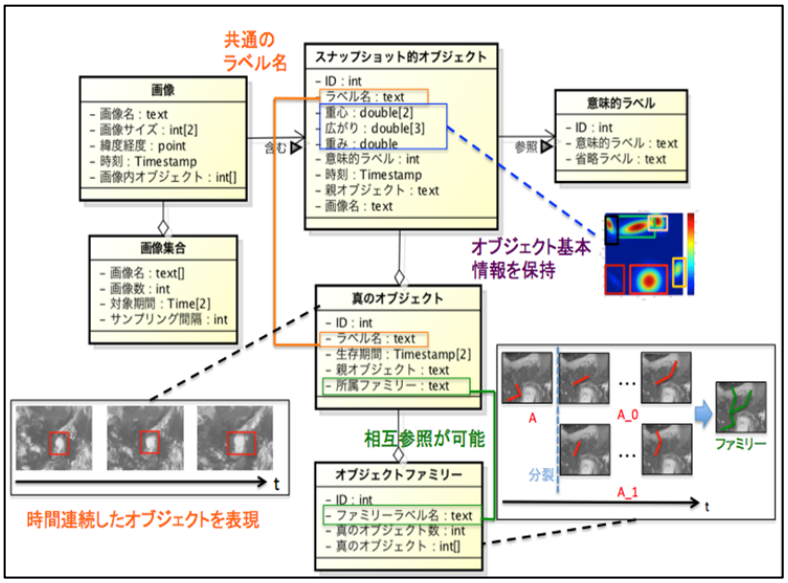


図 7 オブジェクト（ホットスポット）のデータモデル([8]に基づく)

3.3 実装例 -気象画像-

実装例としてここでは 2006 年１年間の日本付近のひまわり画像（MTSAT-7，高知大気象情報頁[9]）に対して構築したシステムの例を紹介する[4]．3 時間おきにサンプリングし，640 x 480 pixel に整形された日本付近の 2920 枚の画像に対してホットスポット抽出をおこなった．ここでは先行研究との用語の統一のために，ホットスポットをオブジェクトという言葉で表現するものとする．ホットスポット（オブジェクト）の意味的ラベリングにはオブジェクトを切り出した画像の FFT パワースペクトルを特徴量として自己組織化マップ[11]を使用してクラスタリングさせ，さらに人間が当時の天気図をもとにしてクラスタごとに意味付けすることによって行った．

実験の結果，抽出されたオブジェクト数は 14058，所要時間は Intel core i5 3.2GHz (iMac)で 7094sec

であった．なお学習に要した時間は同じ環境で 4239sec であった．クラスタリングでは初期に 36 ユニットに分類された結果について，類似のクラスタをスーパークラスタとしてまとめることによって，最終的に 7 種のグループにまとめ，意味的ラベル(台風，台風の一部，寒冷前線，梅雨前線，温暖前線，停滞前線，それ以外)を付与した．意味的ラベルの確からしさをクラスタごとの適合オブジェクトの割合（精度: Precision ）で評価するとその平均値は 61 % 程度であった．この値はパターン認識の精度(最近の畳み込みニューラルネットワーク等による画像分類のコンペティション ILSVRC2016 では上位では 95%以上達成[12]，実用のためには 80%以上程度は必要と考えられる)としては低いが，これは教師なしのアルゴリズムで発見的にクラスタリングを行っていることによるもので，教師学習アルゴリズムを使用することで性能を向上させられると考えられる．

図 8 に構築されたシステムのトップページを示す．月ごと，もしくは年ごとにどのような種類のホットスポットがどれだけ抽出されたのか，その頻度分布が表示されている．さらに時間指定をすることによって，図 9 のようにスナップショット画像の上に抽出したオブジェクトの領域を表示し，その意味的ラベルを色で表示するビューを開くことができる．図 9 の対象画像には２つの台風が含まれているが１つは台風としての認識に成功し，もう１つは失敗していることがわかる．さらにそれぞれのオブジェクトの軌跡を表示することもできる．

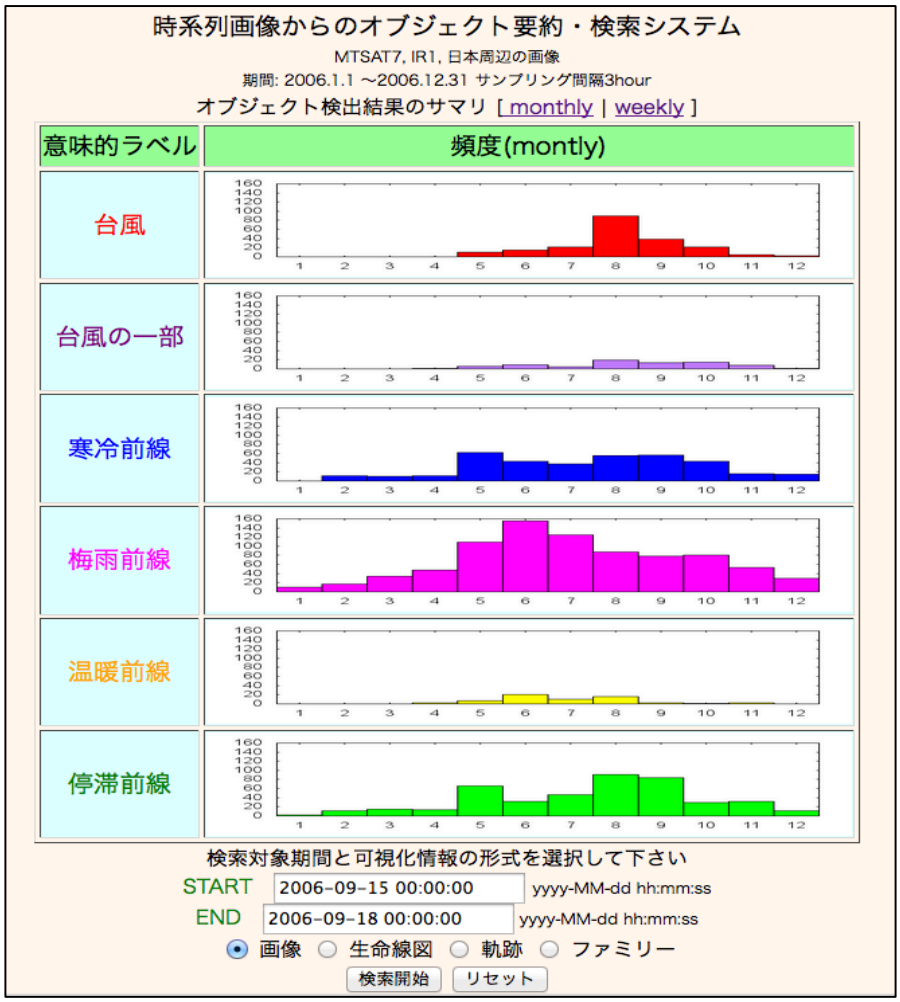
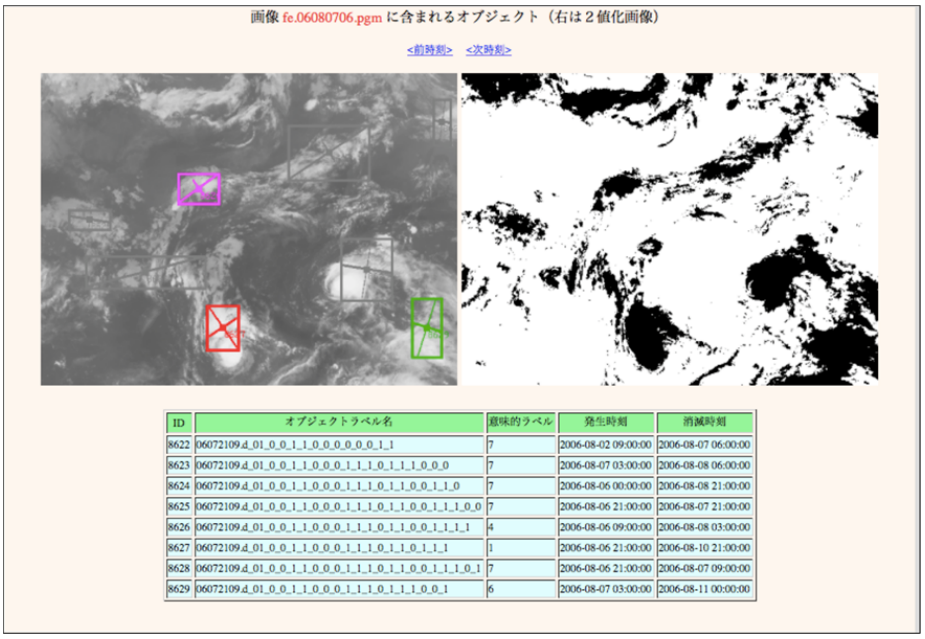


図 8 構築されたシステムのサマリー頁([8]に基づく)

また図 10 のように，横軸を時間，縦軸をオブジェクト ID として，UML の生命線図に類似した形で，オブジェクトの生存期間や分裂，相互作用，共起関係の視覚的な発見に用いることのできる画面を表示することができる．これらは高次の知識発見の前に，ユーザが試行錯誤をしながらインタラクティブにパターン発見を実施し，仮説を検証するのに活用できると考えられる．図 10 を詳細に見ると，同じオブジェクトでも時間によって異なる意味的なラベルがつけられていることがわかる（グラフ上部の赤枠内で台風が頻繁にその他にラベル付けされているなど）．これは，現象としてはホットスポットの性質が時間とともに代わりうると考えられることから自然なことではあるが，本来はある安定度をもって同じラベルが付与されるべきもので頻繁に入れ替わるべき物ではない．これは意味的ラベル付けの精度がまだ良くないことと，各時間で独立にラベル付けをしてしまっていることの影響をうけている．意味的ラベル付けの精度向上とその時間的連続性の考慮は，今後検討すべき課題である．



ID	オブジェクトラベル名	意味的ラベル	発生時刻	消滅時刻
8622	06072109.d_01_0_0_1_1_0_0_0_0_0_0_1_1	7	2006-08-02 09:00:00	2006-08-07 06:00:00
8623	06072109.d_01_0_0_1_1_0_0_0_0_1_1_1_0_1_1_1_0_0_0	7	2006-08-07 03:00:00	2006-08-08 06:00:00
8624	06072109.d_01_0_0_1_1_0_0_0_0_1_1_1_0_1_1_0_0_1_1_0	7	2006-08-06 00:00:00	2006-08-08 21:00:00
8625	06072109.d_01_0_0_1_1_0_0_0_0_1_1_1_0_1_1_0_0_1_1_1_0_0	7	2006-08-06 21:00:00	2006-08-07 21:00:00
8626	06072109.d_01_0_0_1_1_0_0_0_0_1_1_1_0_1_1_0_0_1_1_1_1	4	2006-08-06 09:00:00	2006-08-08 03:00:00
8627	06072109.d_01_0_0_1_1_0_0_0_0_1_1_1_0_1_1_0_1_1_1	1	2006-08-06 21:00:00	2006-08-10 21:00:00
8628	06072109.d_01_0_0_1_1_0_0_0_0_1_1_1_0_1_1_0_0_1_1_0_1	7	2006-08-06 21:00:00	2006-08-07 09:00:00
8629	06072109.d_01_0_0_1_1_0_0_0_0_1_1_1_0_1_1_1_0_0_1	6	2006-08-07 03:00:00	2006-08-11 00:00:00

図9．スナップショットで抽出されたホットスポットの表示．オブジェクトの色は図８の意味的ラベル毎の表示と同様（たとえば，台風は赤）．灰色はその他のオブジェクトに分類されたことを示す．（[8]に基づく）

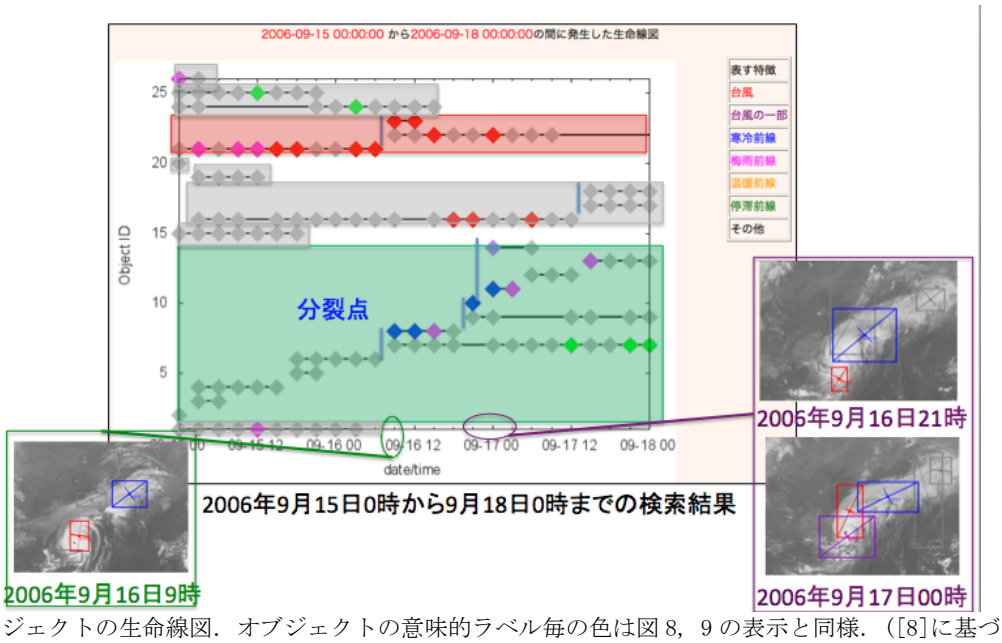


図 10．オブジェクトの生命線図．オブジェクトの意味的ラベル毎の色は図 8，9 の表示と同様．（[8]に基づく）

4. 議論と今後の検討課題

以上の内容から汎用的な時空間グリッドデータからのホットスポット抽出，要約システムの構築の検討課題についてまとめる．気象画像やフェーズドアレイ気象レーダへの適用結果から，フィールド値の値そのものがホットスポットの中心や多変量正規分布の形状を規定している場合とそうでない場合を識別し，その両者に適したアルゴリズムを示した．特に現在検討を始めているフェーズドアレイ気象レーダデータからの降水コアの抽出[10]に関してはフィールド値を活用することによってより精密なモデリングが可能になる可能性が予備的に示された．またスナップショットオブジェクト，真のオブジェクト，オブジェクトファミリーの３つでデータモデルを整理することによって，インタラクティブな知識発見システムを構築することが可能になった．

これらによって，ホットスポットの自動抽出，要約のための基本的な要件が押さえられたと考えられるが，(a) 時間的連続性についての単純な扱い，(b)意味的ラベルの精度の不足などの問題が残っている．これらの問題は互いにかかわり合っているが，意味的ラベルの精度不足については，教師有り学習の活用がその解決策としてあげられる．一方，時間的連続性の取り扱いが単純過ぎることで，新規発生と分裂が混同される，意味的ラベルが頻繁に入れ替わるといった問題が起こっていることに関しては，事後の検証で修正する手法

と，時間変化を考慮したモデルに拡張することが考えられる．短期的な検討としてはまず現在の検討の自然な延長として事後の検証を取り入れることによる修正を検討する予定である．これらを検討しながらリアルデータでの実装システムの構築と，汎用システムの構築を目指した基礎的な検討を並行して進めていく予定である．

5. まとめ

本研究では，時系列からホットスポットをオブジェクトとして抽出，要約することによって時空間知識発見を行うためのシステムを構築するための基礎手法とモデルについて検討し，その結果を気象画像に適用した例を紹介した．現時点で，基礎的な手法と検討が出そろった状態であるが，一方で，(a)時間的連続性の扱い，(b)意味的ラベルの精度の不足などの問題が現れている．今後はこれらの問題についてリアルデータでの実装システムの構築と，汎用システムの構築を目指した基礎的な検討を並行して検討していく．

謝辞

気象衛星画像の解析を担当した松永知也氏，3 次元に対応した手法の開発を分担した松岡愛美氏，フェーズドアレイ気象レーダデータを提供いただいた NICT 佐藤晋介氏，村田健史氏に感謝します．また有益な意見をいただいた査読者に感謝します．

参考文献

[1] 佐藤晋介，牛尾知雄，水谷文彦，フェーズドアレイ気象レーダの研究開発，NICT News，2013.1， 2013，3-5.
[2] MacQueen，J.，Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability (Vol. 1，No. 14)，1967，281-297.
[3] Dempster，A. P.，Laird，N. M.，and Rubin，D. B.，Maximum likelihood from incomplete data via the EM algorithm. Journal of the royal statistical society. Series B (methodological)，1977，1-38.
[4] Bishop，C. M.，Pattern Recognition and Machine Learning，Springer Verlag New York，lLC.，2006.
[5] Schwarz，G.，Estimating the dimension of a model，Annals of Statistics 6，461-464.
[6] Honda，R.，Wang，S.，Kikuchi，T. and Konishi，O.，Mining of objects from time-series images and its application to satellite weather imagery，Journal of Intelligent Information Science，19:1，2002,79-93.
[7] 松永知也，本田理恵，時系列画像からのオブジェクトベースデータマイニング -オブジェクトの抽出とデータベース化-，DEIM フォーラム 2015，2015，P2-4，pp.6.
[8] 松永知也，森啓太，本田理恵，時系列画像に含まれるオブジェクト特徴の変遷要約とその可視化，DEIM フォーラム 2016，2016，P1-2，
[9] 高知大学気象情報頁,http://weather.is. Kochi-u.ac.jp，2017.6.30 参照.
[10] 林諒，本田理恵，佐藤晋介，村田健史，村永和哉，鵜川健太郎，佐々浩司，村田文絵，時系列 3 次元グリッドデータからのホットスポットの自動抽出・追跡法の開発 -フェーズドアレイ気象レーダデータによる局地的大雨解析への適用-，DEIM フォーラム 2017，2017，P4-4，pp.7.
[11] Kohonen，T.，Self Organizing Maps，Springer，3rd ed.，2000.
[12] UNC vision lab，Large Scale Visual Recognition Challenge 2016 (ILSVRC2016)，http://image-net.org/challenges/LSVRC/2016/index，2017.