

# 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 ーオブジェクトの抽出とデータベース化ー

松永 知也<sup>†</sup> 本田 理恵<sup>‡</sup>

<sup>†</sup> 高知大学大学院総合人間自然科学研究科 〒780-8520 高知県高知市曙町 2-5-1

<sup>‡</sup> 高知大学理学部情報科学教室 〒780-8520 高知県高知市曙町 2-5-1

E-mail: <sup>†</sup> matsunag@is.kochi-u.ac.jp, <sup>‡</sup> honda@is.kochi-u.ac.jp

**あらまし** 近年様々な分野で大規模な時系列画像のアーカイブが作成されている。時系列画像には時間とともに形を変えながら移動する対象物“オブジェクト”を含む問題が頻出する。膨大な時系列画像からオブジェクトを抽出し、記述することで、その発達過程や相互作用のパターンといった高次の知識発見を行うことが可能になることが期待される。本研究では、時系列画像から多変量正規分布の混合分布によるモデリングを用いて、時間的特徴変化を考慮しながら半自動的かつ高速にオブジェクトを抽出し、その特徴を記述する手法を提案する。またオブジェクトの誕生、分裂、融合、消滅や相互作用を記述するオブジェクトのデータモデルについても述べ、さらに気象画像への適用例も示す。

**キーワード** 時系列画像, オブジェクト, 多変量正規混合分布, 気象画像

## 1. はじめに

近年様々な分野で大規模な時系列画像のアーカイブが作成されている。時系列画像においては時間とともに変形しながら移動する対象物“オブジェクト”を含む問題が頻出する。こうした時系列画像からオブジェクトを抽出して、その時空間変動パターンを理解するために、データマイニングの手法を適用することが期待される。

時系列画像の主な例としては気象衛星画像や監視カメラの映像などが挙げられる。気象画像ではオブジェクトは雲塊、監視カメラでは人物等の移動体ととらえることができる。また、こうした分野では技術の伸展により、高解像度化、サンプリングレートの増加、多チャンネル化などにより、さらに大量のデータが蓄積されるようになってきている。例えば、2014年10月に打ち上げられたひまわり8号では、7号よりも空間解像度が2倍、サンプリング間隔が1/3になって、よりきめ細かいレベルの気象現象の理解が期待されている<sup>[1]</sup>。

このように大量に蓄積されたデータから時空間変動パターンなどの高次の知識を発見するには、オブジェクトの情報を抽出し、記述することが重要である。膨大な時系列画像からオブジェクトを抽出してその発達過程(気象画像における天候の変化など)や相互作用のパターン(共起など)を求めることができれば、自然・社会現象の動態把握に活用できると期待される。

Honda et. al.(2002)<sup>[2]</sup>では多変量正規分布の混合分布を用いて不特定数・不特定形状のオブジェクトの抽出をおこない、気象画像を例にその手法の有用性を確認した。また、石津(2009)<sup>[3]</sup>では、時系列画像における時間連続性を考慮して、2枚目以降の入力画像に対し、

前の時間の解を初期値として与え、その成分数に揺らぎをもたせてモデルパラメータの推定をおこなう手法を提案した。これにより、オブジェクト抽出処理の時間短縮、同一オブジェクトの追跡、分裂するオブジェクトの親子関係(どのオブジェクトから派生したか)を明示することを可能にした。しかし、この2つの研究ではオブジェクトの抽出手法しか提案されておらず、オブジェクトの重要な情報である特徴(台風、前線など)の記述はおこなわれなかった。

本研究では、時系列画像から多変量正規混合分布によるモデリングを用いて時間発展の履歴を保持しながら半自動的かつ高速にオブジェクトを抽出するとともに、その時間的特徴変化を記述する手法を提案する。さらに得られたオブジェクトの情報をデータベース化することで抽出された情報を視覚的容易にとらえられるようにし、また高次の知識発見支援を目的とした発達過程の可視化についても検討する。最終的にはユーザーがWebブラウザを通してオブジェクト情報を検索し、データベース内の情報や可視化画像の取得が可能な知識発見支援システムの構築を目指す。そのために必要な作業であるオブジェクトの誕生、分裂、融合、消滅や相互作用を記述するオブジェクトのデータモデルについて検討し、さらに気象画像への適用例も示す。

第2章では、オブジェクト抽出、意味的ラベリング、オブジェクトの発達の記述の手法について述べる。第3章では構築する時系列画像からの知識発見支援システムの案を示す。第4章では、入力された画像における前処理、オブジェクト抽出、分類、記述までの実験と実験結果について述べる。第5章では得られたオブジェクト情報をデータベース化する際に必要となるデ

ータモデルについて検討する．最後に全体のまとめと今後の課題について述べる．

## 2. 手法

### 2.1. 想定するシステムの概要

図1に想定するシステムの概要を示す．まず時系列画像集合からオブジェクトを抽出して，その形状，位置を記録し，テキストチャからラベリングを行う．さらに前後の親子関係から，分裂や融合を反映した族（ファミリー）の情報を付加して，オブジェクトのデータベースに格納する．この情報からさらに共起性や時間依存ルール，軌跡のクラスタリング等のより高次の知識を発見してデータベースに格納し，webサーバを通じてユーザーが変動パターンを抽出・可視化し，この過程を繰り返すことによってインタラクティブなパターン発見等を行うものとする．本発表では，このシステムの要素技術として，オブジェクト抽出，ラベリング，族を含む情報の表現，生命線の可視化を説明する．

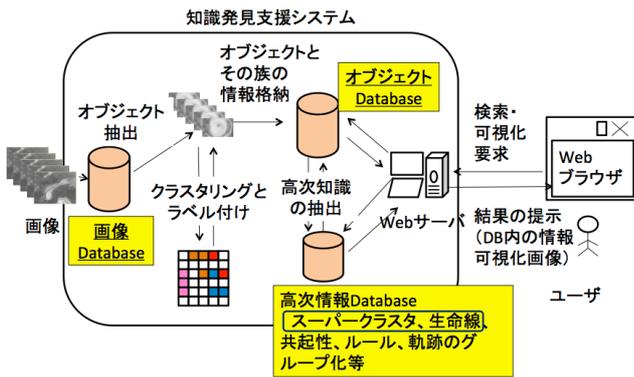


図1. 想定する知識発見システムの概要

### 2.2 オブジェクトのモデル

図2にオブジェクト抽出の概念図を示す．時系列画像に含まれるオブジェクトを抽出する際には，前処理として，入力された画像に対して2値化・サンプリングをおこなう．この2値化画像のサンプリング点  $X_i = (x_i, y_i)^T$  がオブジェクトの存在を表すものとする．

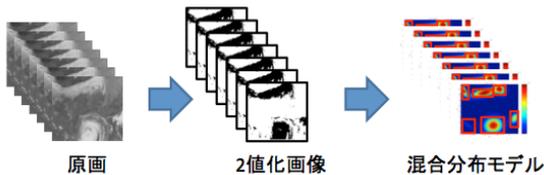


図2. 多変量混合分布によるオブジェクトのモデリング

オブジェクト  $j$  の分布は多変量正規分布により以下の通り表すものとする．

$$p(X|\mu_j, S_j) = \frac{1}{2\pi |S_j|^{1/2}} \exp\left\{-\frac{1}{2}(X-\mu_j)^T S_j^{-1}(X-\mu_j)\right\} \quad (1)$$

ここで  $X = (x, y)^T$  はオブジェクトの存在する点， $\mu_j = (\mu_{xj}, \mu_{yj})^T$  は平均値(オブジェクト  $j$  の中心)， $S_j$  は分散共分散行列(オブジェクトの広がり)を表す．

$$S_j = \begin{bmatrix} \sigma_{xxj} & \sigma_{xyj} \\ \sigma_{yxj} & \sigma_{yyj} \end{bmatrix}. \quad (2)$$

複数のオブジェクトの分布は正規分布の混合分布で表現する．

$$P(X|\theta) = \sum_{j=1}^O \omega_j p(X|\mu_j, S_j), \quad (3)$$

$$\sum_{j=1}^O \omega_j = 1, \quad (4)$$

$$\theta = \{(\omega_j, \mu_j, S_j) | j = 1, 2, \dots, O\}. \quad (5)$$

ここで， $\omega_j$  は重み係数， $O$  はオブジェクトの個数， $\theta$  はモデルパラメータである．

### 2.3 オブジェクト抽出

先行研究<sup>[2][3]</sup>に基づくオブジェクト抽出の前処理は次の通りである．

まず，原画像を2値化してオブジェクトの存在する座標をサンプリングし，EMアルゴリズム<sup>[4]</sup>によってあるオブジェクト数に対して対数尤度を最大化するパラメータを求める．この際，最適解が得られることは保証されないので，初期値をランダムにばらつかせて，複数回(10-20回)の実行結果から対数尤度が最大の結果を選択する．

なお，含まれるオブジェクト数(正規分布の成分数)  $O$  についてはありうる値の範囲について総当たりで推定し，ベイズ情報量基準  $BIC$ <sup>[5]</sup> に従って最適な成分数を選択する．

$$BIC = -2\log L(\theta) + (6o - 1)\log N \quad (6)$$

ここで  $L(\theta)$  は尤度， $N$  はデータの個数である． $BIC$  を最小にする成分数  $o$  を選択することによって，オブジェクトの個数も自動決定出来る．

また，オブジェクトは時間的に連続にその位置，形状を変化させることから，前の時間の解を次の時間の解の初期値として利用することで，解の収束の安定化が期待出来る．またこのことによって，オブジェクトの追跡や，消滅，融合の検知も可能となる．

図3にこの手法の概念図を示す．1枚目の画像では，あり得る成分数に対して総当たりで  $k$ -means法<sup>[6]</sup> で初期値を与えてEMアルゴリズムで解をもとめ，その解から  $BIC$  を最大化する解を選択する．2枚目以降の画像では直前の画像のモデルパラメータを基にして， $\pm 2$  の範囲で成分数にゆらぎをもたせた解を初期値とする．成分数削減の際には，重み係数  $\omega_j$  の小さい成分から順

番に削減し、一方、増加させる場合には、重み係数 $\omega_j$ の大きい成分を2重化させた。また、解には初めて出現した時間のラベルを継承して付けた。これらの処置によって計算時間の短縮とともに、同一オブジェクトの追跡や子オブジェクトへの分岐の過程を記述することを可能にする。

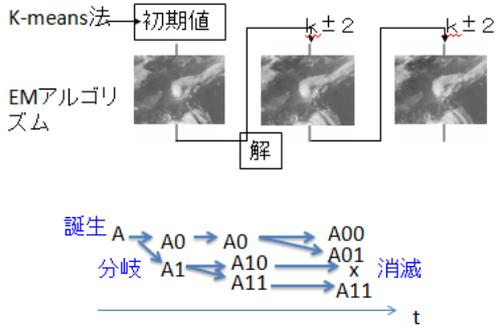


図 3. 時間連続性を考慮したオブジェクト抽出

## 2.4 意味的ラベリング

さらに抽出されたオブジェクトに対して意味的ラベリングをおこなう。図 4 にその流れを示す。まず、オブジェクトの重心と広がりをもとに最小包含正方形領域(内接正方領域と外接正方領域)をオブジェクト画像として取り出してスケールリングする。抽出されたスケールリング画像に対して、オブジェクトのテクスチャと形状を表す特徴ベクトルとして、高速フーリエ変換(FFT)パワースペクトルを求める。次に学習用にサンプリングした特徴ベクトルのデータセットに対して自己組織化マップ(self organizing map, SOM<sup>[7]</sup>)によるクラスタリングをおこなう。これにより空間パターンが似た特徴をもつオブジェクトが自己組織化マップ上の近いユニットにクラスタリングされる。得られた各クラス集合に対して、その内容を手作業で吟味して、台風、前線等の意味的ラベルの付与をおこなう。この結果を全オブジェクトに適用することで、全オブジェクトに意味的ラベルを付加する。

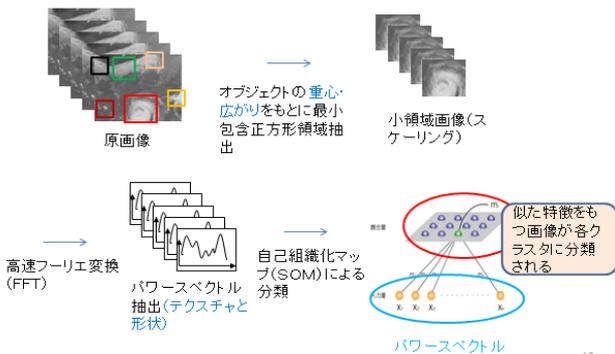


図 4. 抽出されたオブジェクトのクラスタリング過程

## 3. 実験

### 3.1 実験条件

実験には気象庁から提供されて高知大学気象情報頁<sup>[8]</sup>に蓄積された気象衛星ひまわり 7号 IR1(10.3~11.3 $\mu$ m)で撮影された日本付近(北緯 70 度から南緯 20 度, 東経 70 度から東経 160 度の領域)の画像を利用した。サンプリング間隔は 6 時間とし、2013 年の 1 月, 6 月, 10 月の各 2 週間分, 計 168 枚を使用した。各月の典型的な画像を図 5 に示す。

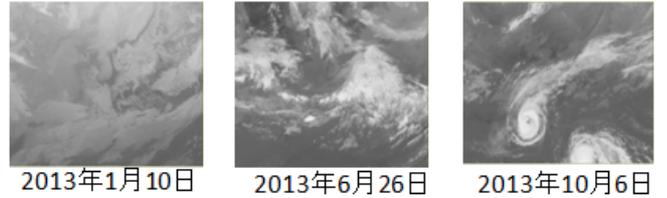


図 5. 使用した気象画像例(各月の典型的な画像を示す)

オブジェクト抽出実験では、初期成分数を 3~18 とし、各 10 回の試行で BIC によって最適モデルを求めた。自己組織化マップによるクラスタリングでは、抽出した 1662 個のオブジェクトの内、矩形領域内に画像の欠けがない 768 個を学習に用いた。自己組織化マップの出力層は 6x6 ユニットとし、学習時の近傍距離 2, 学習係数 0.3, 学習回数 10000 回とした。

意味的ラベル付けには図 6 に示す気象庁発表の天気図<sup>[9]</sup>も参照した。オブジェクトにラベル付けをおこなう際には原画像と同じ日付の天気図を比べてその位置関係などから意味的ラベルを決定した。

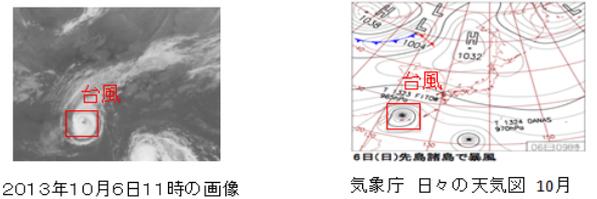


図 6. ラベル付けの例(赤枠は台風).右図は気象庁発表の天気図<sup>[8]</sup>.

### 3.2 結果と考察

図 7 にオブジェクトの抽出結果例を示す。決定された成分数は順に 8, 9, 9 で、その混合分布のみカラーカウンターで示している。ここで楕円状の成分がそれぞれのオブジェクトに相当する。一例として、マゼンタの最小包含矩形で示されたオブジェクトが変化しつつ・移動している様子を追跡できていることがわかる。また、1 枚目に存在していた赤枠で囲まれたオブジェクトが 2 枚目で消滅していることや、2 枚目の画像で新しく黄色の枠で囲まれたオブジェクトが誕生(分裂)

したことも正しくとらえていることが分かる。

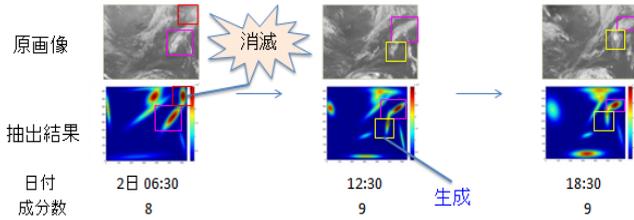


図 7. 抽出結果の例(10月2日6時30分から連続した3枚)

図 8 には取りだしたオブジェクトの自己組織化マップによるクラスタリングの結果を各クラスタの代表画像(図 8 左)と、それぞれに付与した意味的ラベル(図 8 右)として示す. 同じような特徴をもつ画像が競合層上で近くに配置されていることが分かる. 各クラスタは台風, 梅雨前線, 冬型の雲などに特徴付けられる. さらに同種のクラスタを結合して意味的なスーパークラスタを形成する. 図 8 右で同色のクラスタが1つのスーパークラスタを意味する. ここでは, 5 種類のスーパークラスタ, 台風, 台風の一部, 梅雨前線, 寒冷前線, その他を設定した. ”その他”には, 冬期の雲など特徴のあるものも含まれるが, 本研究ではオブジェクトに注目して時空間変動パターンをとらえようとしているため, 特に塊状のオブジェクトが含まれない(もしくはオブジェクトの全体像が認識しにくい)クラスタについては1つにまとめた.

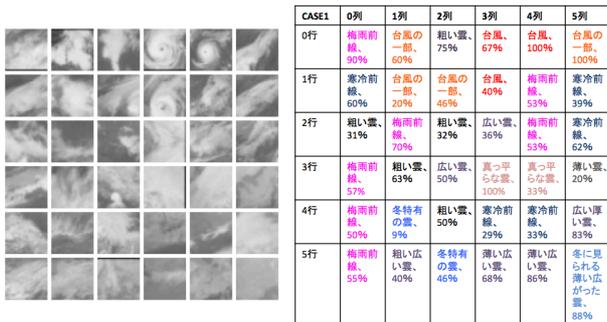


図 8. 自己組織化マップによる特徴分類の結果(数字は適合率)

クラスタリングの精度を評価するためにクラスタ毎, および全体の平均適合率を求めた (図 8 にラベル下の数字で表示). 各クラスタの分類精度(適合率)は同一クラスタ内の適合画像の割合として求めた. 各クラスタの精度は 20-100 %, 全体での平均値は 61.4 %という結果になった. 今回の結果では温帯低気圧が台風や台風の一部として分類されており, オブジェクトの形状だ

けで特徴を分類するのは難しいことがわかる. また, 同様に温暖前線と寒冷前線の判別も難しい. 今後は複数のバンド幅を利用するなどして分類結果の向上を目指す必要がある.

これらの予備実験により, 高次の知識を発見するためのオブジェクトの基礎データを半自動的に収集できるようになったことを確認した.

### 3.3 高次のパターン

次に, 抽出された情報から 2 次的に得られる親子関係や軌跡について紹介する.

図 9 にオブジェクトの中心軌跡を可視化した 3D 鳥瞰図(横軸 x 座標, 縦軸 y 座標, 高さ時間軸)を示す. ここでは異なる色の軌跡はランダムに異なるオブジェクト毎に割り振っているが, 親子関係で色を統一し, 選択的に表現することで, その分岐や移動パターンを直感的にとらえることができると思われる.

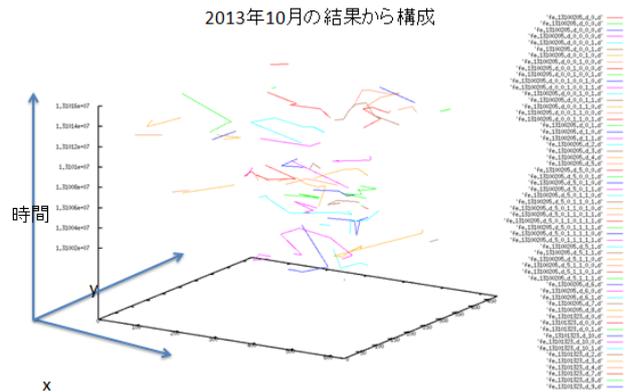


図 9. オブジェクト軌跡の可視化例(2013年10月)

図 10 にはオブジェクトの生命線図を示す. ここでは横軸が時間を表し, 縦軸はオブジェクト ID を示す. データ点は赤色が台風, オレンジが台風の一部, ピンクが梅雨前線, 青色が寒冷前線, 黒がその他を表している. また, 各点は時間断面でのオブジェクトを表すようになっている. 例えば図 10 の四角の枠に囲まれている領域(4日0時から10日0時)ではオブジェクトが多数分岐していることから1つの族が表現されていることが分かる. 実際に画像で確認してみると, 画像内に台風の一部が現れ, やがて分裂, 消滅していくのが確認できる. また枠内の赤丸に着目してみると, 台風と前線が共起していることがわかる. このような共起性は時空間変動パターンの理解においては重要な要素であるので, 可視化, 検索, 定量的評価などを含め, 今後検討していく必要がある.

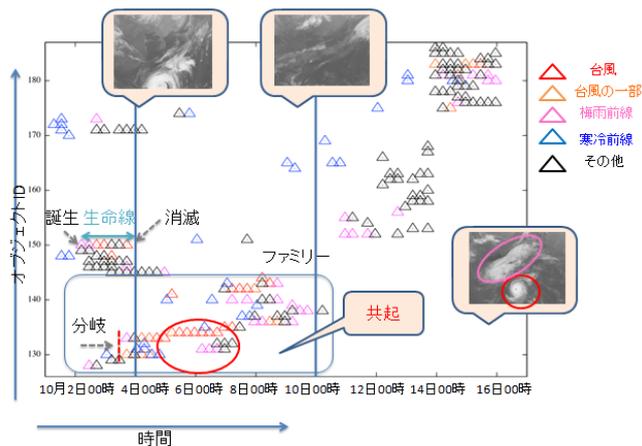


図 10. オブジェクトの生命線図の例(2013年10月)

#### 4. 概念データモデル

得られたオブジェクト情報をデータベース化し、高次の知識発見に利用するためにその概念データモデルを検討する。

この問題において考えられる要素は次の6つとし、各要素の関係を図 11 にクラス図として示す。

- (1) 画像集合(ImageSet)
- (2) 画像(Image)
- (3) 時間断面(画像内)のオブジェクト(Object)
- (4) 複数時間にわたって存在するオブジェクト(ObjectTrajectory)
- (5) オブジェクトの族(ObjectFamily)
- (6) 意味的ラベル(SemanticLabel)

本研究で用いるオブジェクトという用語は、ある時間断面で表されるオブジェクト(Object)と、複数時間にわたって存在するオブジェクト(ObjectTrajectory)の2つをさしている場合がある。実世界での現象を考えると、後者が本来のオブジェクトであるといえるが、いずれの視点も知識発見に必要な情報であるため、別々のデータモデルとして表現することにする。また、オブジェクトの族(ファミリー)も同様に階層的なオブジェクトの概念として扱う。

オブジェクトのデータモデルが持つ属性として、発生時刻と消滅時刻、発生元オブジェクト(親オブジェクト)を記述することで、誕生、消滅、分裂を表現する(融合は消滅の一部ととらえる)。時間断面のオブジェクトはその重心、広がりなど基本的な情報を属性として持つ。オブジェクトの軌跡は時間方向の配列で表現するとともに、検索を容易にするためにその形状は2次元の最小包含矩形(minimum bounding rectangle, MBR)として定義する。オブジェクトの族は属性としてその族

を構成するオブジェクトを含み、一方、オブジェクト自身もどの族に所属しているかの情報を持つものとする。

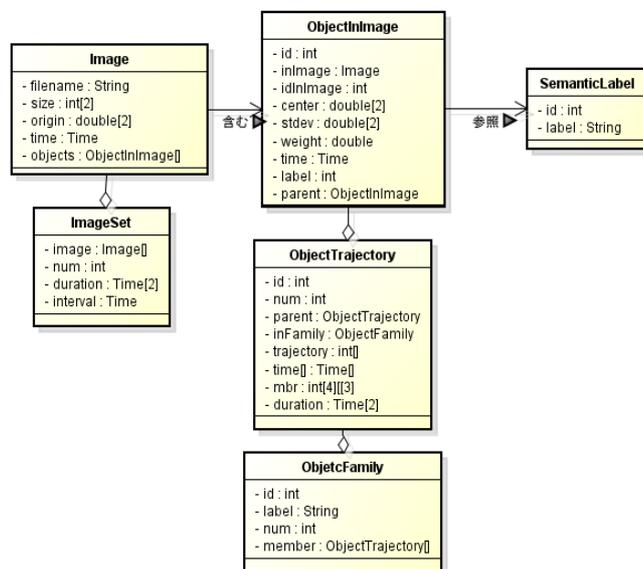


図 11. データモデルの相互関係

なお、時空間データベースのための概念モデルに関する関連研究としては、関本、柴崎による FEO モデル<sup>[10]</sup>があり、ここではモデルの構成要素はフィーチャー、イベント、オブザベーションに整理されている。本研究でまとめた要素と比較すると、(1)画像集合、(2)画像はオブザベーション、(2)画像内オブジェクト、(3)複数時間にわたるオブジェクト、(4)オブジェクト族はフィーチャーに対応し、(3)、(4)についてはイベント的な要素も持つ。特に(4)は親子関係をもって分岐して発達するのが特徴である((5)は FEO には現れない補助的な要素)。これらのモデルでの可能な範囲での表現、拡張の必要性についても今後検討していく予定である。

#### 5. おわりに

本研究では、時系列画像から抽出されるオブジェクトの情報をデータベースに格納し、新たに高次の情報を作成することでユーザーが Web から検索、結果の表示が得られる知識発見支援システムの構築を目指して、時間連続性を考慮したオブジェクト抽出、自己組織化マップによるクラスタリングと意味的ラベルの付与、時間的特徴変化の可視化、データモデルの検討をおこなった。

気象画像へ適用した結果、オブジェクト抽出では、オブジェクトの分裂、消滅といった変化を捉えることが可能になり、また複数時間にわたっての追跡が可能となったことを確認した。ただし、現在の手法では想定外に多数のオブジェクトが同時誕生する等の急激な

変化が起こったときには対応できない。また、大きいオブジェクトと小さいオブジェクトを同時に取り出すことは難しいと予想される。このような場合には異常な状態が発生したことを検知する判断基準や階層的な取り扱いの検討が必要である。また、融合と消滅が同様に扱われ、融合する前のオブジェクトの情報が保持されないことについても改良が必要である。

自己組織化マップによるクラスタリングと意味的ラベル付けにおいてはオブジェクトの形状から台風、梅雨前線といったグループ分けと意味付けが可能になった。ただし意味的ラベル付けには手作業が必要であり、その精度も平均 61.4%と低いため、Deep Learning<sup>[11]</sup>などで教師有り学習を導入することも検討の必要があると考えられる。また、温帯低気圧の分類や前線の区別といったオブジェクトの形状だけでの分類では難しい問題も発見された。これについては形状と組み合わせる利用できる気圧等の他の情報の利用検討が必要と考えられる。

抽出した情報の可視化では、オブジェクトの分裂や消滅の過程を要約表示できることを確認した。また、関連するデータの概念データモデルを検討し、各要素の相互関係やオブジェクトの発生、分裂、融合、消滅を表現することができた。

今後は知識発見支援システムの完成のため、上記の問題の解決とともに、データベースの実装、高次の情報作成の処理ツールの作成をおこない、システムを完成していく予定である。

## 謝辞

有益なコメントをいただいた法政大学三浦孝夫先生他の方々に感謝する。

## 参 考 文 献

- [1] 佐々木, 操野, 静止地球環境観測衛星ひまわり 8号及び9号について, 日本リモートセンシング学会誌, vol 31, no. 2, 255-257, 2011.
- [2] 石津, 時間連続性を考慮した混合分布推定によるオブジェクト追跡—時系列気象画像への適用—, 高知大学卒業論文, 2009.
- [3] R. Honda et al., Mining of objects from time-series images and its application to satellite weather imagery, Journal of Intelligent Information Science, 19:1, pp. 79-93, 2002.
- [4] A. P. Dempster et al., Maximum Likelihood from Incomplete Data via EM Algorithm, Journal of the Royal Statistical Society, series B, vol. 39, no. 1, 1-38, 1977.
- [5] G. Schwarz, Estimating the Dimension of a Model, Annals of Statistics, 6, 461-464, 1978.
- [6] J. B. MacQueen, Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability 1, University of California Press, 281-297, 1967.
- [7] T. Kohonen, Self Organizing Maps, Springer, 3rd ed., 2000.
- [8] 高知大学気象情報頁, <http://weather.is.kochi-u.ac.jp>.
- [9] 気象庁, 日々の天気図 <http://www.data.jma.go.jp/fcd/yocho/hibiten/>.
- [10] 関本義秀, 柴崎亮介, 時空間データベースのダイナミックな更新を目指した概念データモデルの提案, GIS-理論と応用, Vol.8(1), pp.63-73, 2000.
- [11] 麻生秀樹, 多層ニューラルネットワークによる深層表現の学習, 人工知能学会誌 28 卷 4 号, 2013.