

知識発見支援のための気象画像データベースの構築

片山 幸治[†] 小西 修^{††}

画像データベースにおけるデータマイニングは、大規模な画像データの中に存在している自明でない関係やパターンなどを発見することを目的としており、科学的発見法とも言われている。本稿では、膨大な気象画像データに対して、そのデジタル画像から気象の情報を取り出し、画像データベースに変換し、専門家による解析と知識発見を支援することを目的とした気象画像データベースについて述べる。そこでは、雲の濃淡情報とその変化からなる画像データに、我々が開発したクラスタリングと相関ルールといった二つのデータマイニング法が適用される。これによって、パターン化された気象画像の時系列データや年間において約 24 種類の気象の特徴パターンが得られる。これらの画像データは、R-Tree を用いた画像データベースとして構築され、専門家の“春頃で、前線の後に移動性高気圧が来るような気象変化を見たい”のような要求に応えることができる。

Construction of Satellite Image Databases for Supporting Knowledge Discovery

KOJI KATAYAMA[†] and OSAMU KONISHI^{††}

Image data mining deals with the extraction of implicit knowledge, image relationship, and patterns not explicitly stored in image databases. It belongs to the scientific discovery. In this paper, We focus weather satellite images with many image information and describe on the construction of weather image databases that can search patterns like a series of cloud or change of weather condition. we apply the algorithms which we call co-occurrence rules to data mining. It consists of the combination of Self-Organizing feature's Map and association rules. The results take time sequence data of cloud patterns and twenty-four feature patterns extracted over one year. The image databases have been constructed using R-tree and can answer to expert's questions such that we would find any weather patterns which are a migratory anticyclone after a frontogenesis.

1. はじめに

近年、データベースやインターネット技術の進歩に伴って、画像、ビデオ、音声、テキスト情報などの様々なマルチメディア情報の利用が可能となってきた。現在では日本の静止軌道気象衛星ひまわり (GMS-5) により、雲の動きや風の状況、海面温度などの様々なデータがリアルタイムで送られてくる。このような膨大な気象データは気象庁など数多くのウェブサイトで公開されており、過去の気象状況について調べたい場合にも、インターネット等を利用して、正確でより多くのデータ収集が可能である。さらにこれらの気象情報は日々刻々と蓄積されており、これらの膨大なデー

タの中には、有用な情報がたくさん隠されている可能性がある。

このような大量の情報から有用な情報を引き出す手法にデータマイニングがある。データマイニング (Data Mining) は、データベースからの知識発見 (Knowledge Discovery in Databases : KDD) とも呼ばれ、データベースに蓄積された大量のデータから、自明でない規則性やパターンを半自動的に取り出す方法に関する科学研究である¹⁾²⁾³⁾⁴⁾。

ここで巨大科学データからのデータマイニング事例として、SKICAT システム⁵⁾ が知られている。このシステムでは、大規模な天文学的調査によって得られた膨大な天体データに対して決定木による分類を行ない、天文学者による科学的発見を支援する。また、マルチメディア情報におけるデータマイニング事例として、MultiMedia Miner⁶⁾ がある。これは、リレーショナルデータベースからのデータマイニングシステムである DBMiner とデジタルライブラリなどのコンテン

[†] 現在、(株)富士通高知システムエンジニアリング
Presently with Fujitsu Kochi Systems Engineering

^{††} 高知大学理学部数理情報科学科
Dept. of Information Science, Faculty of Science, Kochi University

ツベースの画像検索システムである C-BIRD という 2つの技術に基づいて、大きなデータベースから高いレベルのマルチメディア情報や知識を発見する。このシステムでは、マルチメディアデータを様々な次元から解析するのを促進する Data Cube の組み立てを行ない、画像またはビデオデータベースから要約、分類、相関などの様々な知識の発見を行なう。特に医療画像や衛星画像、ビデオ映像、監視カメラなどの膨大な量の画像情報は数テラバイトにもなり、これらのデータを人手によって詳しく調べるのは手間がかかり現実的ではない。このように画像データベースにおけるデータマイニングは、巨大な画像データの中に存在している自明でない関係や、パターンなどを発見することを目的としており、科学的発見の部類に位置する。

高知大学気象情報頁[☆]には、東京大学生産技術研究所において受信された GMS-5 による観測データが蓄積されており、大規模な気象画像データが存在する⁷⁾。しかし、気象画像はいわばカオス現象とも考えることができ、対象の位置や形状が頻繁に変化するという特徴がある。また一つ一つは単なる画像データであるが、動画のように時間的連続性を持ったデータであり、このような特徴を持つ膨大な画像データに対してコンテンツベースの画像検索を行なうには従来の画像処理技術では限界がある。そこで本研究ではコンテンツベースの画像検索にデータマイニングの手法を適用する。これによりデジタル画像から気象の情報を取り出し、画像データベースに変換して抽出したパターン情報の視覚化を行ない、専門家による解析と科学的発見を支援する気象画像データベースの構築を行なうことを目的としている。

本研究では PGM 形式（画像サイズは 640×480 ）の濃淡画像約 1000 枚を対象として気象画像データからのデータマイニングを行なった。そこでは、画像データを $N \times M$ 個のブロックに分割して、そのブロックの中に出現する雲の濃淡情報のヒストグラムに基づいたクラスタリングを行う。次に気象の推移を画像の特徴となるクラスタにより表現し、時系列データへと変換する。そして得られた時系列データにある種の相関ルールを適用して気象画像の時系列変化の特徴をパターン化する。

例えば、検索例として、“10 月ごろの台風の移り変わりをみたい。”といったものや、“春頃で、前線の後に移動性高気圧がくるような天気図をみたい。”などの要求に応えることができる。

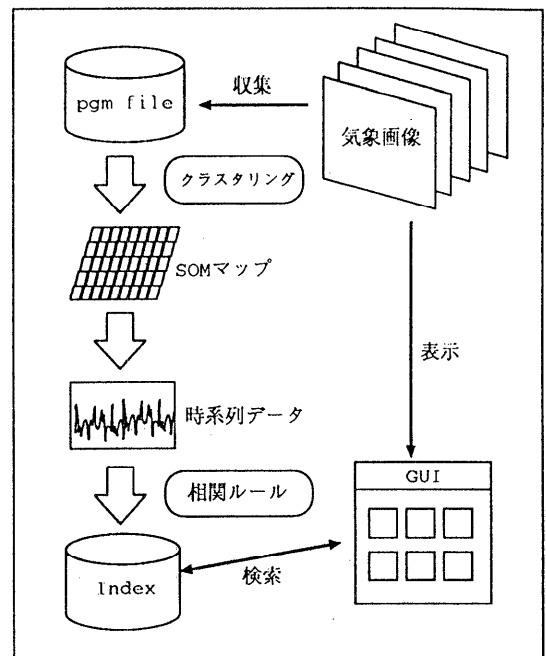


図 1 システムの概要

Fig. 1 The architecture of the system.

本研究の特徴を以下にあげる。

- 画像検索にデータマイニングの手法を適用したこと
- 時系列のデータを対象に特徴を表現したこと
- 特徴ベクトルではなく、分類結果を検索条件インタフェースにしたこと
- 専門家の分析を支援するシステムに組み上げたこと

以下、2 章では知識発見支援システムの概要について、3 章ではクラスタリングによる気象画像からの時系列データ生成について、4 章では時系列データにおける気象パターンの抽出方法について、5 章で構築した気象画像データベースについて述べ、6 章で結論および今後の課題について述べる。

2. 知識発見支援システムの概要

我々は、気象衛星画像からこれらが持つ雲の濃淡情報や、気象の移り変わりなどの変化を持つパターン情報に対して、クラスタリングと相関ルールの二つのデータマイニング法を適用することにより、気象衛星画像の時系列変化を画像データベースとして蓄積し、専門家による解析と知識発見を支援する気象画像データベースの構築を行っている⁸⁾⁹⁾。本システムの概要を図 1 に示す。本システムにおける処理の流れは以下のようになる。

[☆] 高知大学気象情報頁 (<http://weather.is.kochi-u.ac.jp/>)

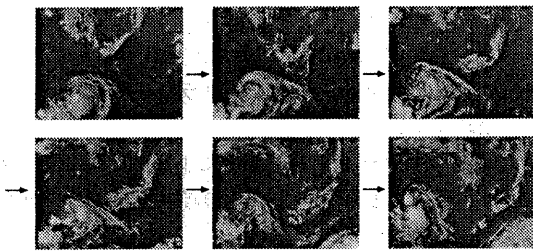


図 2 連続した気象衛星画像の例
Fig. 2 Examples of satellite images.

- (1) 高知大学気象情報頁で提供されている PGM 形式の赤外線画像の自動収集
- (2) 自己組織化マップによる気象画像情報のクラスタリング
- (3) クラスタリングで得られたクラスタをもとに時系列データを生成
- (4) 時系列データに相関ルールを適用して気象の推移などのパターン情報を抽出
- (5) 気象の推移を表わす時系列データにおけるパターン情報の索引付け
- (6) 索引付けされたパターン情報による時系列データの検索と動画表示

以上の過程により、本システムでは一定時間間隔で得られる膨大な気象衛星画像データの半自動的な特徴付けを行ない、気象の推移に基づいた画像のコンテンツ検索を行なうことが可能である。

3. クラスタリングによる気象衛星画像からの時系列データ生成

まず気象衛星画像に対して雲の濃淡情報によるクラスタリングを行なう。これにより各画像の特徴付けを行ない、クラスタの特徴で表現する。次に気象の推移を各画像の特徴の流れで表わされる時系列データに変換する。

3.1 気象衛星画像の特徴とクラスタリング

高知大学気象情報頁では、東京大学生産技術研究所において受信された GMS-5 による観測データの利用が可能であり、最新の画像データに対してアクセスを行なうことができる。また図 2 に示すように本研究で使用了画像は PGM 形式の日本列島上空における雲の濃淡画像 (640 × 480) であり、毎日 1 時間ごとのデータが収集され、年間で約 9000 枚の大規模な過去の画像データが蓄積されている。

このような膨大な気象衛星画像に対してそれらが持つ雲の特徴等のコンテンツ情報を用いて検索を行なう場合に、気象衛星画像の特徴として、

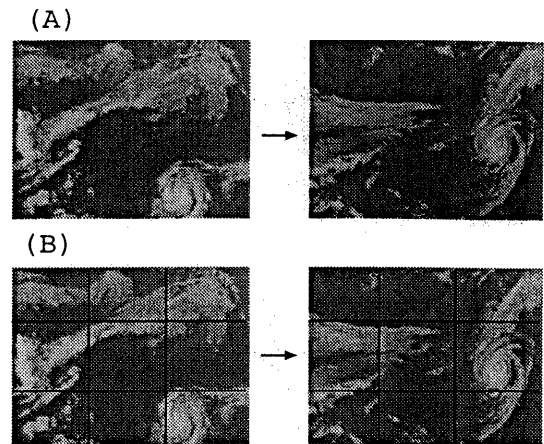


図 3 クラスタリングの問題点
Fig. 3 Problems of clustering.

- (1) 雲の位置や形状といったように頻繁に変化を伴う画像データである。
- (2) ビデオ映像のように時間的連続性を持った時系列データである。

などの理由から、このように頻繁に雲の形状が変化し、位置が定まっていない画像データに対して、画像の一部や特徴などを検出するのは従来の画像処理技術では限界がある。そこで本研究では、雲の濃淡情報に基づいて分類を行なうために、Kohonen の自己組織化マップ (Self-Organizing Map : 以下 SOM)¹⁰⁾ を適用し、気象画像データのクラスタリングを行なう。SOM は、1990 年に T.Kohonen によって提案されたパラダイムであり、ベクトルで表わされる入力パターン間の位相関係を、学習アルゴリズムにより発見、分類して位相地図を組織化する 2 層のネットワークである。

しかし、画像データに対してクラスタリングを行なう場合に画像のピクセルをそのまま SOM の入力ベクトルとすると、図 3 の (A) に示すように同じ対象であっても異なるクラスタに分類されてしまうといった問題が起きる。そこで図の (B) に示すように画像をブロックに分割してクラスタリングを行なうことにより、そのブロックの中に出現する雲の特徴をとらえる。このようにクラスタリングを行なうことによって、移動する対象であっても同じクラスタに判別させることができる。

3.2 自己組織化マップによるクラスタリング

本研究におけるクラスタリングの手順は次のようになる。まず、画像を $N \times M$ のブロックに分割してそのブロックの中に現われる雲の特徴を SOM によって特徴付けする (これをパターンセルと呼ぶ)。しかしここで、画像の分割における大きさをどのように設定す

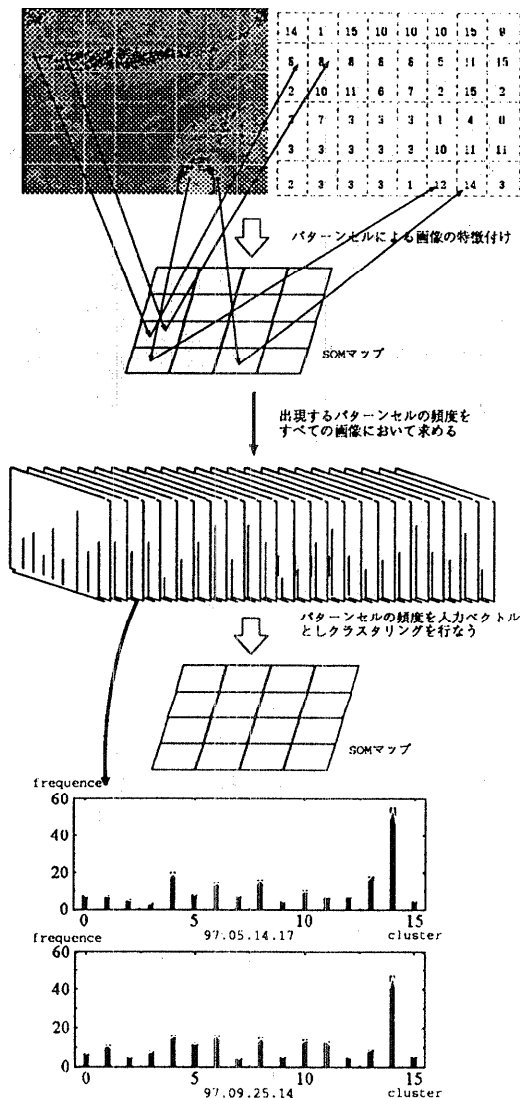


図4 SOMによる気象衛星画像のクラスタリング
Fig. 4 Clustering satellite images by SOM.

るかによってクラスタリングの結果が変わる可能性がある。そこで本手法ではブロックの分割数はできるだけ大きくし(6×8または12×16), 1回目のSOMによって特徴付けられたパターンセルの頻度情報から得られるヒストグラムをSOMの入力ベクトルとして, 2回目のクラスタリングを行なう。これにより雲の濃淡情報の出現頻度に基づいたクラスタリングを行う。クラスタリングを行なう手順をまとめると以下のようになる(図4参照)。

第1段階 パターンセルのクラスタリング

step1 全ての画像をN×Mのブロックに分割して, N×M×画像枚数のブロックを生成する

step2 ブロックのピクセルを入力ベクトルとし学習を行わせ, SOMマップの生成を行う。

step3 元画像をSOMマップクラスタのアドレスで表現する。

第2段階 パターンセルの頻度によるクラスタリング

step1 出現するパターンセルの頻度を全ての画像ごとに求める

step2 各画像のパターンセルの出現頻度によるヒストグラムをSOMマップの入力ベクトルとし, 学習を行う。

本研究では, 1997年のPGM形式の赤外線画像から, サンプルングとして8時間ごとの1日3枚のデータを取り出し, 1044枚を対象として実験を行なった。1段階の画像分割におけるブロック数N×Mは196(12×16)とし, SOMマップは1段階, 2段階ともに4×4のグリッドとしてクラスタリングを行なった。図4に示しているパターンセルの頻度の例では, 5月と9月といったように異なった季節に存在する画像であっても類似した雲の出現頻度であれば, 同じクラスタにクラスタリングされる。

そこで, このクラスタリングによる分類能力がどの程度の精度を持っているのかを判断するために, SOMによってクラスタリングされたクラスタの精度を, 「適合率」と「再現率」によって求める。またこの分類の精度が高ければ, 4章で述べる時系列データからのパターン抽出における精度は高くなる。ここで, クラスタに入らなかった適合画像をA, クラスタに入った適合画像をB, クラスタに入った不適合画像をCとしたとき,

$$\text{適合率} = \frac{B}{B+C}, \quad \text{再現率} = \frac{B}{A+B}$$

と表わされる。クラスタリングを行った気象衛星画像1044枚全体の中で, それぞれのクラスタにおいて, クラスタに入らなかった適合画像は合計146枚, クラスタに入った不適合画像は合計168枚であった。この結果を表1に示す。本手法では正確な雲の位置情報を扱っているわけではないので, 出現頻度が類似していても同じ雲の特徴をとらえていることにはならないが, 表1の結果が示すように, 今回の実験では比較的良好な結果が得られている。

表1 クラスタリングの精度
Table 1 Recall and precision

適合率	再現率
86.0(876/1022)	84.6(876/1044)

Landscape: クラスタの特徴の説明		
ファイル	編集	表示 ジャンプ Communicator
クラスタ	季節	目立つ特徴
00	春夏	前線、台風
01	春秋	西高東低
02	春秋	西高東低
03	春秋	帯状の高気圧
04	秋	移動性高気圧
05	春秋	前線
06	秋から冬	筋状の雲
07	春秋	帯状の高気圧
08	夏	太平洋高気圧、前線
09	春から夏	梅雨前線、台風
10	冬	冬型、渦状の雲
11	冬	冬型、渦状の雲
12	夏	太平洋高気圧、台風
13	春夏	高気圧、台風
14	冬	寒冷前線
15	春、秋	移動性高気圧

図 5 各クラスタの特徴付け例

Fig. 5 Examples of characterized cluster.

また、クラスタの専門的な特徴付けを行うために、これらの各クラスタに対して月別頻度を求め、それをもとに各クラスタを季節別に分類した(図5)。そして図に示すように、それぞれのクラスタに見られる前線や台風などの特徴を視覚的に判断し、クラスタの専門的な特徴付けを行なった。この過程において画像を分割するブロックの大きさやSOMにおけるグリッド数などの変更を専門家と協議して行なうことにより、より細かい雲の特徴付けを行なうことが可能である。

3.3 時系列データの生成

クラスタリングを行い、専門的に特徴付けを行なうことによって、各画像データをそれらの特徴別に半自

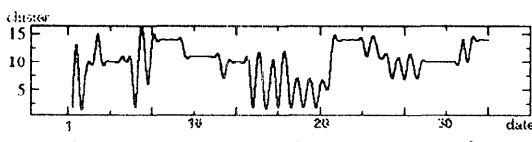


図 6 約 30 日間のクラスタの移り変わり

Fig. 6 The changing patterns of clusters over one month.

動的に分類することが可能になる。ここで、気象衛星画像の特性として、図2で示したように時間的連続性を持った時系列データであるということがあげられる。この特性に注目して、各気象画像に割り当てられた標本となるクラスタの数値を時間軸上に順に列挙する。これを図6に示す。これは約30日間のクラスタの移り変わりを示したものであり、縦軸はクラスタ番号、横軸は日付を表している。

このようにすべての画像データを全体的な流れとしてとらえた場合には、特徴のあるパターンが隠されている可能性が考えられる。ここで、同じクラスタが連続して出現する場所や、同じようなクラスタの移り変わりを持つ場所は、全体の流れの中で特徴をもったパターンである。この流れの中に成立している気象画像データ間の規則性や、関係を導き出すことにより、大量の気象画像データの中に存在するパターンを導出する。これは温度変化や、地震波形、音声情報といったようなパターンと同様である。

4. 時系列データにおける気象パターンの抽出

気象の推移を表わす時系列データから、変化する雲のパターン情報を抽出するために、相関ルール的一种である共出現対 Term Pairs を抽出する手法と気象の変化における前後関係を抽出する手法について述べる。

4.1 気象の時系列データにおける共出現対の抽出

相関ルール¹⁾(Agrawal, et.al., 1993)は、データベースの関係が与えられたときに、あるタプルにおいてある支持度(support)を満たす条件 X に対して、ある確信度(confidence)を越えた確率で、条件 Y が満たされるというルールのことで $X \Rightarrow Y$ と表現される。相関ルールの抽出は通常以下の2つのステップに分けることができる。

- (1) 与えられたタプル集合から、最小支持度を満足するアイテム集合をすべて求める。
- (2) 上のステップで求めたラージアイテム集合から、最小確信度を満たすすべてのルールを見つけ出す。

我々は、共出現対の頻度情報による順序関係を用いて、相関ルールを抽出する共出現ルール法を開発してきた¹¹⁾(小西, 1989)¹²⁾。この手法では、大規模な(約10万件)文献情報を対象に、文献の中に共出現する重要ターム(キーワード)間の情報に注目した共出現対 Term Pairs の抽出を行なう。そしてこの共出現対集合を自己組織化マップによってクラスタリングし、そのクラスから共出現ルールを抽出することによって、与えられた文献集合の特徴を表す概念マップを創出で

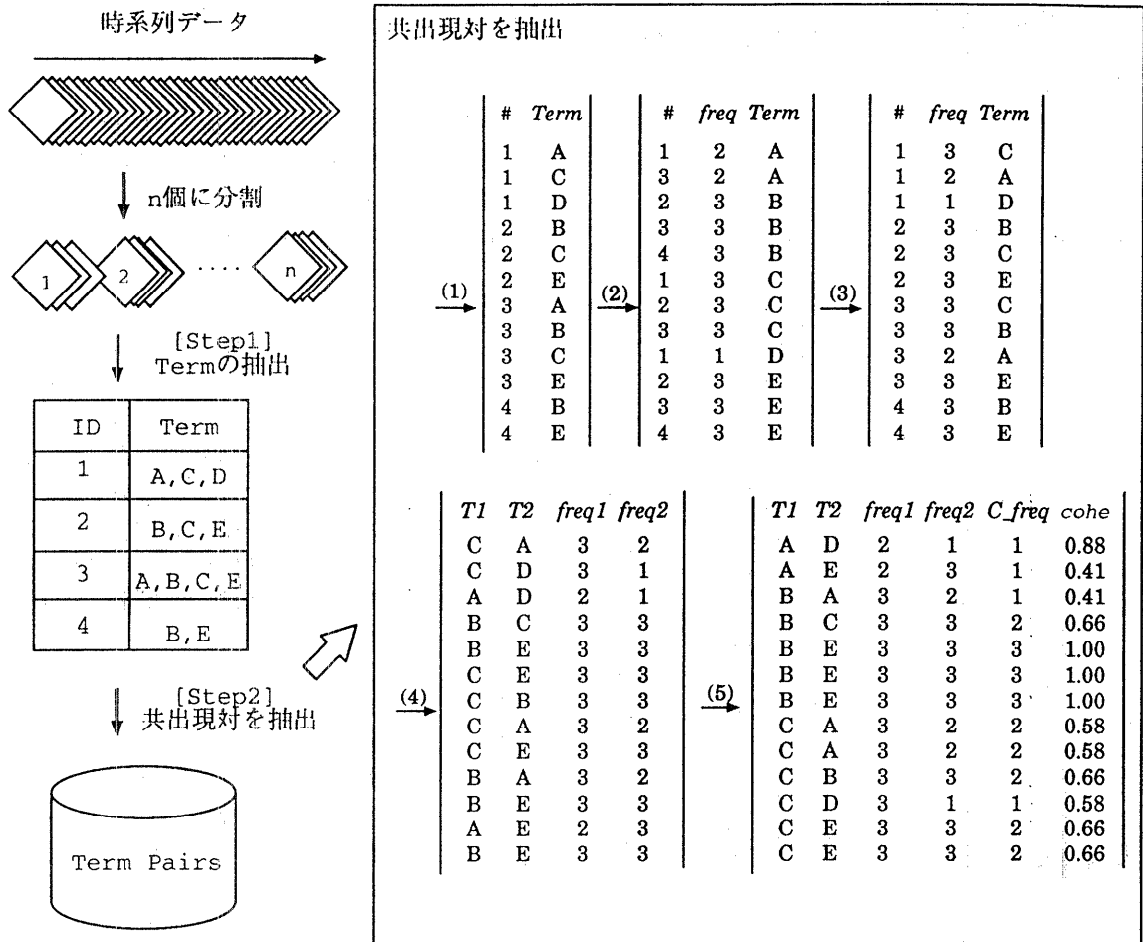


図 7 時系列データにおける共出現関係の抽出過程

Fig. 7 Extraction steps of co-occurrence relationship.

きるものである。

以下では、本研究で対象とした気象の推移を表わす時系列データを対象として、共出現対を抽出するアルゴリズムについての説明を行なう（以下の説明については図 7 を参照）。

気象の推移をビデオ映像としてとらえると、クラス番号の推移により構成される時系列データはそれぞれ数日間の気象の推移の集合を表わすことになる。そこでまず、時系列データを n 個のセグメントに等分割する（今回の実験では年間の各月における上/中/下旬の 36 個に分割した）。ここでこれらのセグメント集合はビデオ映像におけるシーンまたはショットと考えることができる。そこで各セグメント集合において出現頻度の高いクラスを Term とし、次の Step1 により抽出する。

Step1 Term の抽出

- (1) セグメント集合から抽出したクラスターの頻

度統計をとる

- (2) あるしきい値以上のものを Term リストとする

次に、各セグメント中の Term 間の共起関係の情報を得るために同じセグメント中に共出現する Term と Term の対（共出現対）を求める。ここで Term 間の共起関係を定義する共出現関係について説明する。

定義 1 セグメント集合 S を $S = (S_1, S_2, \dots, S_n)$ とする。セグメント集合から抽出されるクラスターを $TERM_k = (t_{1k}, t_{2k}, \dots, t_{nk})$, (t_{ik} はセグメント S_i の Term) とすると、共出現対は $TermPair(TERM_k, TERM_h)$

$$= \{[t_{ik}, t_{ih}] | t_{ik} \in S_i, t_{ih} \in S_i\}$$

となる。

定義 2 $TermPair(TERM_k, TERM_h)$ に重みを付けるために、 $TERM_k$ と $TERM_h$ の間の距離（結合度）を次のような関数で与える。

$Cohesion(TERM_k, TERM_h)$

$$= \frac{freq.ofTermPair(TERM_k, TERM_h)}{\sqrt{[freq.ofTERM_k \times freq.ofTERM_h]}}$$

ここで, $freq.ofTERM_k = \sum t_{ik}$,

$$freq.ofTERM_h = \sum t_{ih}$$

このような共出現対の二項関係を共出現関係と呼ぶ。この共出現関係を次の Step2 により各セグメント集合から抽出し、共起して出現するクラスタの共出現対を取り出す（各クラスタの順序関係は考慮しない）。これは図 7 の右側に示しており、Step2 の (1) から (5) は図 7 の (1) から (5) に対応する。

Step2 共出現関係の抽出

- (1) すべての Term を列挙する
- (2) Term の出現頻度を求める
- (3) 同じ ID で頻度の降順にソートする
- (4) 同じ ID においてペアを作る（組み合わせをとる）
- (5) 共出現するペアの頻度と結合度を求める

以上のステップにより、 n 個に分割したセグメント集合において共起して出現するクラスタの二項関係を抽出する。これにより与えられた時系列データから同じようなクラスタの出現度をもつセグメント集合を求める。今回の実験では年間を 36 個に分割し、年間の各月における上/中/下旬を対象として類似パターンの抽出を行なった。同様の手法で、時系列データにおける分割数を増やすことにより、年間において 1 日または数日ごとの気象の類似パターンを求めることができる。

4.2 連続したクラスタの前後関係に注目したパターン抽出

気象の推移を表わす時系列データにおいて、気象パターン A の後に気象パターン B が見られるような前後関係を反映した気象の推移となる近接対を抽出する手法について述べる（図 8 参照）。まず、時系列データ（図 8 の (A)）に対して、連続して出現するクラスタを同じ気象パターンをもつ連続した画像の特徴とみなして、ある期間における特徴の系列とする。このとき連続するクラスタを SC (Series Cluster) と呼ぶこととし、 $A' = SC(A, Sf, Ts, Te)$ と表現する。ここで A はクラスタ、Sf はクラスタの連続度、Ts, Te は Term の開始および終了時刻である。図 8 における A' , B' , C' は連続度 Sf を 2 以上とした場合である。

ここで、 B' , A' , C' , A' , B' という並びから順に ($B'-A'$, $B'-C'$)、($A'-C'$, $A'-B'$)、... といったように近接する SC の対を抽出していく（図 8 の (B)）。この例では次の SC とその次の SC との組み合わせを抽出している。次に近接した SC の結合度を与えるために共

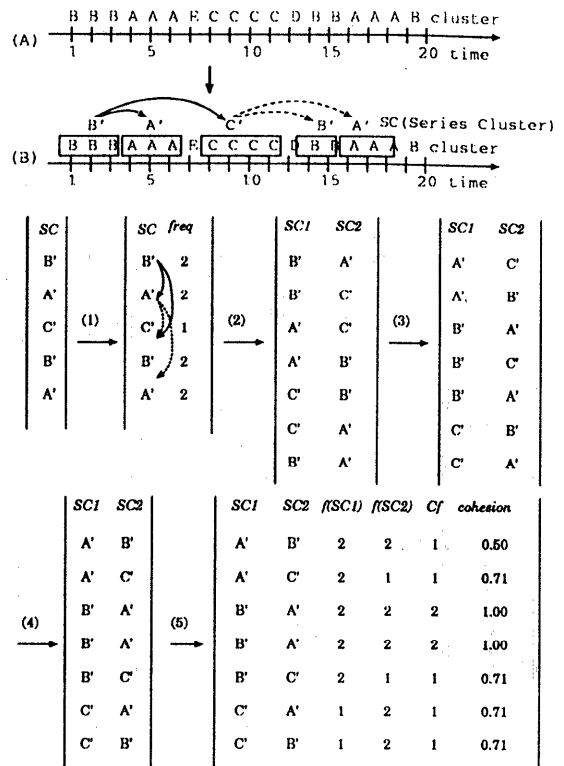


図 8 特徴パターンの抽出

Fig. 8 Extraction of feature patterns.

出現関係における定義 2 の式を適用する。以下にその過程を示す。ここで各数字は図 8 の各数字と対応している。

- (1) 各 SC の出現頻度 $freq$ を求める。
- (2) 近傍距離のしきい値を定めて近接する SC の組み合わせを取り出す。
- (3),(4) 辞書式順序でソートする（SC1, SC2 は近接する SC）。
- (5) 近接する SC の組み合わせの頻度 Cf と結合度 $cohesion$ を求める。

このように連続したクラスタに着目することにより、同じクラスタに分類された画像を同じ気象パターンをもつ連続した画像とみなして、時系列データから前後関係を反映した気象の変化パターンとなる近接対を抽出する。これによりある気象パターン A の後にある気象パターン B が見られるといったように前後関係を反映した気象の推移を抽出する。

5. 気象画像データベース

我々は、以上に述べたデータマイニング法を用いて抽出したパターン情報を索引情報としてデータベース



図 9 検索画面
Fig. 9 Search form.

に格納し、専門家による解析と科学的発見を支援する気象画像データベースの構築を行なった。

5.1 気象画像データベースの構成

本システムでは、

- (1) “10月ごろの台風の移り変わりをみたい”
- (2) “今日の雲のパターンと同じようなパターンを示す過去の日をみつけない”
- (3) “春頃で、前線の後に移動性高気圧がくるような気象変化をみたい”
- (4) “冬から春にかけて、渦状の雲の後に南からの高気圧が見られるような気象変化がみたい”

といったような質問に答えることが可能である。(1), (2)の検索要求に対しては、4.1節で抽出した共出現対を検索キーとし、クラスタリングによる分類結果を検索条件インターフェースとして類似検索を行なう。次に(3), (4)は時系列データに基づく前後関係を考慮にいたれた検索要求であり、これらの検索要求に対しては4.2節で抽出した近接対にR-tree¹³⁾を用いてPostgreSQL上で索引付けを行なった。また本システムの特徴は、静止画像の類似検索を行なうだけでなく、動画データベースとして気象衛星画像の検索を行なうことができる。現時点では、動画の表示にJavaアプレットを使用している。

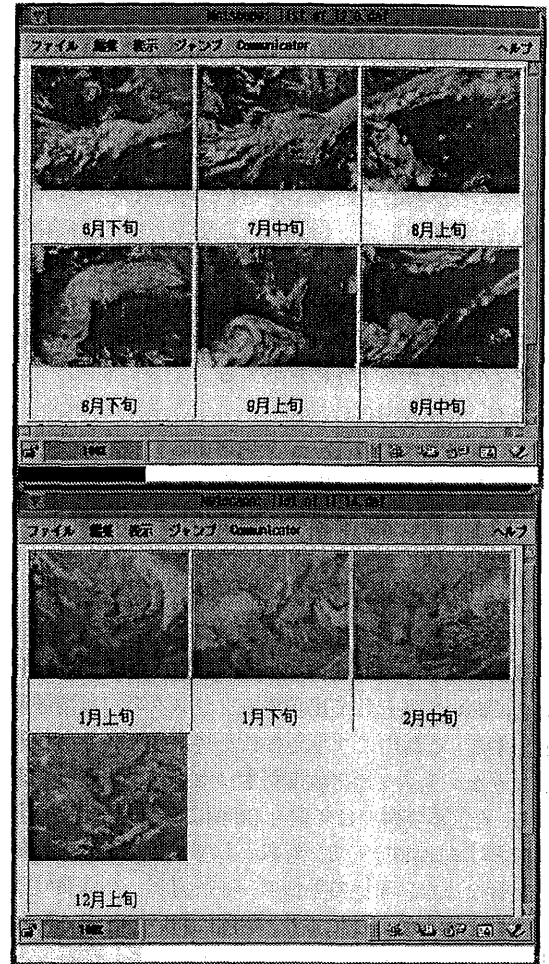


図 10 共出現対 12-0 と 11-14 による検索結果
Fig. 10 An example of results by term-pair(12-0 and 11-14).

5.2 類似検索

検索要求(1), (2)における類似検索について述べる。図9に類似検索の検索画面を示す。図の右のフレームはSOMによってクラスタリングされたセル数16個のマップを表わしており、同じクラスタにクラスタリングされた画像の平均値をとり、それに一番近い画像をそのクラスタを代表する索引としている。また図の左上に専門的にクラスタの特徴付けを行なった例を示しており、左下のフレームに時系列データから得られた共出現対のデータを示している。このシステムでは、抽出された共出現対 TermParis を検索キーとして2つのクラスタを指定し、各セグメントの類似検索を行う。ここで、左下のフレームの T1, T2 は Term, f1, f2 は各 Term の頻度, Cf は共出現の頻度, cohesion は共出現の結合度を表わす。cohesion のしきい値を0.5以上とした場合、14個のパターンが抽出された。

* <http://sun.is.kochi-u.ac.jp/>で公開中。

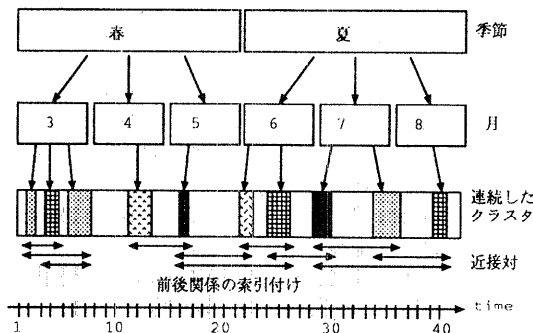


図 11 R-tree による索引付け
Fig. 11 R-tree index.

クラスタリングされたデータにおいて顕著な特徴を持つ 10 個のクラスと合計すると年間において約 24 個のパターンが得られている。

この抽出した共出現対のうちで、クラス 12 と 0 によって類似検索を行なった例を図 10 の上側に、またクラス 11 と 14 によって行なった例を図の下側に示す。この結果、共出現対 12-0 では夏から秋にかけての前線や台風が、また、共出現対 11-14 では冬における渦状の雲が検索結果として得られている。

5.3 時系列検索

次に検索要求 (3), (4) に対する前後関係を考慮にいたれた検索について述べる。4.2 節で得られた特徴の系列となる近接対の開始値と終了値となる画像の ID に対して R-tree を適用し、時系列データの索引付けを行なう。作成方法を図 11 に示す。図に示すように、春は 3 月から 5 月を包囲しており、同様に夏は 6 月から 8 月を包囲している。まずこの関係に注目して、季節と月の包囲関係を索引付けする。そして近接対の開始値と終了値を索引付けすることにより季節または月に包囲される近接対はその包囲関係をキーとして索引が構築される。今回の実験では時系列データから抽出した 116 個の近接対の索引付けを行なった。

この R-tree による索引から近接対によって、時系列検索を行なうシステムを図 12 に示す。このシステムでは図の上のフレームで季節(または月)と前後の気象パターンを選択することにより検索を行なう。この例では、「夏ごろで梅雨前線のあとに台風がみられる天気図」の検索を行なっている。この結果、図の左下に夏における 7 つの期間が検索結果として得られた。この検索結果から 1 つを選択したものが、図の右下に動画像として表示される。

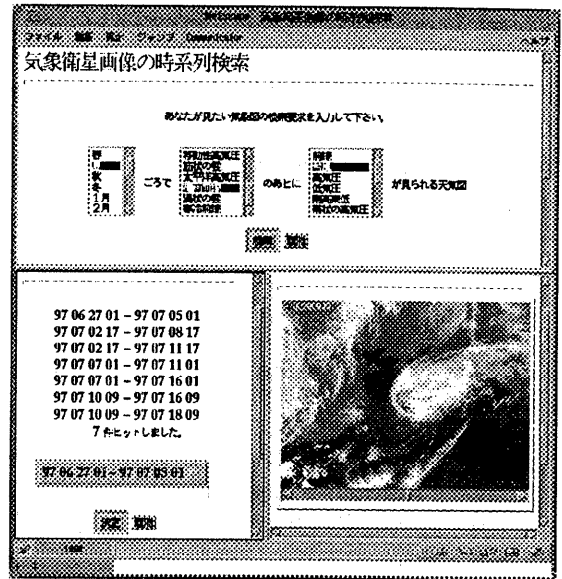


図 12 時系列検索の例
Fig. 12 Search results of sequential data.

6. おわりに

膨大な気象画像データに含まれる雲の濃淡情報とその変化データに対して、クラスタリング(自己組織化マップ法)と相関ルール(共出現関係)という基本的なデータマイニング法を組み合わせることにより、気象画像の時系列変化の特徴のパターン化を行なった。これにより抽出したパターン情報を画像データベースとして蓄積し、専門家による解析と知識発見を支援可能とするシステムを構築することによって、本研究で開発したデータマイニング法の有用性を確認した。

今後の課題として、次の 2 つがあげられる。

- (1) 従来の内容ベースの画像検索は、コンテンツの特徴ベクトルを計算して置き、検索時に参照する方法を用いているが、本研究では、特徴量を自動分類し、分類結果に基づいて、コンテンツ検索を行なうことができる方法を与えているため、膨大なデータにも適用でき、また、過去のデータや新しく蓄積されるデータにも適用できる可能性がある。他の科学データや画像データに適用することで、本手法の有用性を検証しコンテンツベース画像検索技術の高度化に貢献すること。
- (2) 時系列データに対して、ウィンドウを持つ相関ルールを生成し、事象の特徴パターンを抽出するという本研究の手法は、ビデオ画像、音声情報のような連続パターンを持つ様々なデータに対する幅広い適用可能性を持っている。本手法の動画像、

音声情報への適用したシステムの有効性を確認すること。

謝辞 気象衛星画像の利用にあたって、ご協力頂いた高知大学助教授菊地時夫氏に感謝いたします。なお、本研究の一部は科学研究費特定領域研究「高度データベース」による。

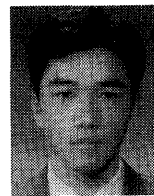
参 考 文 献

- 1) R. Agrawal, T. Imielinski and A. Swami : Mining in Association Rules Between Sets of Items in Large Database. , Proc. ACM SIGMOD Int. Conf. Management of Data , 1003 , pp207-216(1993).
- 2) R. Agrawal , R. Srikant : Fast Algorithms for Mining Association Rules , Preceedings of 20th International Conference on VLDB , pp.487-499 (1994).
- 3) Fomg , J.(Edt.) : "Data Mining, Data Warehousing & Client / Server Databases" , Proc. 8th Int. Database Workshop , Springer (1997).
- 4) Alex, A.F. , and Simon, H.L. : Mining Very Large Databases with Parallel Processing , Luwer Academic Publishers (1998).
- 5) Fayyad, U.M. , Djorgovski , S.G. , and Weir, N. : Automatic teh Analysis and Cataloging of Sky Surveys , In Advances in Knowledge Discovery and Data Mining , pp.471-493 , AAAI Press/MIT Press (1996).
- 6) O. R. Zaiane , J. Han , Z. N. Li , J. Y. Chiang , and S. Chee : MultiMedia-Miner : A System Prototype for MultiMedia Data Mining , Proc. ACM-SIGMOD Conf. on Management of Data , (system demo) , Seattle, Washington (1998).
- 7) 菊地時夫 , "気象衛星 GMS 画像の処理とデータベース化体制の確立による長時間の熱収支の解明 (序報)" , Mem.Fac.Sci.Kochi Univ. (Inform.Sci) , 19 , March (1998)
- 8) 片山幸治 , 小西 修 , 大規模気象画像データベースにおけるデータマイニング法 , 信学技報 , Vol.98 , No.486 , pp33-40 , DE98-30 , (1998)
- 9) 仲川亜希 , 片山幸治 , 金山智一 , 小西 修 , 菊地時夫 : 画像データベースのためのデータマイニング法の拡張 , 情処研報 , Vol.98 , No.34 , pp71-78 , 98-DBS-115 , (1998)
- 10) Kohonen. T. , Self - Organizing Maps , Springer , 1995.
- 11) 小西 修 : 自動構築型知識に基づく専門用語形成システム , 情報処理学会論文誌 , Vol.30 , No.2 , pp179-189 , (1989)
- 12) 仲川亜希 , 小西 修 : 情報探索のための自己組織化アプローチ , 情処研報 , Vol.96 , No.103 , pp39-46 , 96-DBS-110 , (1996).
- 13) Guttman A. : R-tree: A Dynamic Index Structure for Spatial Searching , Proc. ACM SIGMOD Int. Conf. Management of Data , pp.47-57 (1994).

(担当編集委員 河野 浩之)

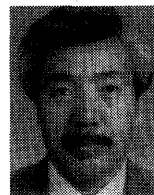
(平成 10 年 12 月 20 日受付)

(平成 11 年 3 月 27 日採録)



片山 幸治 (学生会員)

1974 年生. 1999 年高知大学大学院理学研究科情報科学専攻修士課程修了. 同年より (株) 富士通高知システムエンジニアリング入社.



小西 修 (正会員)

1942 年生. 高知大学理学部数理情報科学科教授. 高知大学情報処理センター長 (1999 年 4 月から). 京都大学工学博士. データベース, 情報検索, 協調情報システムに関する研究に従事. ACM, 電子情報通信学会各会員