

データマイニング法による映像情報の内容検索

金山 智一, 瀧本 裕一, 小西 修
高知大学理学部数理情報科学科
780-8520 高知県高知市曙町 2-5-1
{kanayama, takimoto, konishi}@is.kochi-u.ac.jp

概要

我々は以前に、大規模な気象画像データにクラスタリング（自己組織化マップ）と相関ルール（エピソードルール）というデータマイニング法を組み合わせ、順序関係を反映した特徴パターンの抽出を行う手法を提案した。これは、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。今回、我々は、この手法を野球中継の動画像に適用し、“先行のチームのバッターがヒットを打ったシーン”などのようなシーンが検索可能な動画像データベースの構築を行った。これによって野球中継で試合を構成する主なイベントを抽出でき、試合中で起こる一般的なシーンの抽出が行える。また、特徴パターンの抽出により有用な情報が得られることから情報圧縮の方面での適用も考えられる。

キーワード：データマイニング, クラスタリング, 自己組織化マップ, 相関ルール, エピソードルール, 動画像データベース

Content-based Retrieval of video databases by Data Mining Methods

Tomokazu KANAYAMA, Hirokazu TAKIMOTO, Osamu KONISHI
Dept. of Information Science, Faculty of Science, Kochi University
2-5-1 Akebono-cho Kochi 780-8520 Japan
{kanayama, takimoto, konishi}@is.kochi-u.ac.jp

Abstract

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 video data (a broadcast of a baseball game) with many image information and describe on the construction of video databases that can search patterns like a series of scene or change of scene 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 events and feature patterns extracted over one game. The video databases have been constructed and can answer to user's questions such that we would find any scene as "the batter of the team of precedence hit".

Key words: data mining, clustering, self-organizing map, association rule, episode rule, video database

1 はじめに

近年、WWW(World Wide Web)の普及、データベースやデジタル技術の進歩に伴い大量の様々な情報がデジタル情報に変換され蓄積されており、これらの情報の有効利用が求められている。有効利用を行なうために“データマイニング”や“データベースからの知識発見”という研究が盛んになっており実用化も始まっている。

動画像情報は様々な情報を含んでおり非常に利用価値が高く、効果的に検索利用できる要求が高まっている。現在実用化されているのは、人手により場面場面に内容記述を付けているものである。しかし、人手による内容記述だけでは限界がある。動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実用的でない。そこでこのコンテンツベースの画像検索にデータマイニング [1][2] を適用する。本研究ではクラスタリングと相関ルール [3] を組み合わせることによって半自動的に処理を行なう。

我々は、すでに気象画像で実験を行なった [9][10]。気象画像は、時間的連続性を持った時系列データであり同じ手法で動画像にも適応できると考えた。本研究では、画面の構図が簡単で、同じようなシーンが繰り返し現われ、結果が分かりやすいと考え、プロ野球中継の動画像を実験対象とした。

動画像は静止画像の時系列データであると考え、動画像を連続した静止画像に変換し実験を行なった。まず、静止画像をクラスタリングし、投球シーンなどのコンテンツ付けを行ない、その時系列データに対して、時間依存の相関ルールであるエピソードルール [5] を適用することにより時系列の中にある変化のパターン情報を抽出する。

以下、2章では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時系列データの生成について、4章では動画像の時系列データにおけるエピソードルールの抽出について、5章でシーン検索について述べ、実際の検索例を示す。最後に6章で本論文のまとめおよび今後の仮題について述べる。

2 動画像データベース

我々は、動画像の時系列変化を画像データベースとして蓄積し、解析と知識発見を支援する動画

像データベースの構築を行なっている。本システムでは、対象となる動画像を連続した静止画像に変換し、それをクラスタリングし、そのクラスタを元に静止画像の特徴付けを行ない、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の索引付けを行なう。本システムにおける処理の流れは以下ようになる。(図1)

1. 自己組織化マップによるクラスタリング
2. 得られたクラスタから時系列データを生成
3. 時系列データからエピソードルールを抽出
4. 時系列データの動画像検索

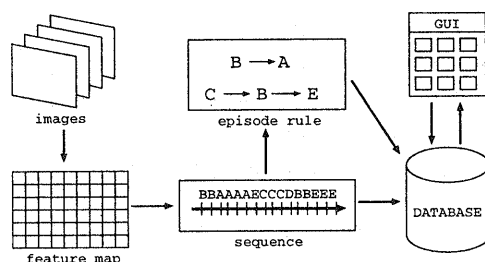


図 1: システムの概要

以上の過程により、本システムでは連続した静止画像の半自動的に特徴付けを行ない、エピソードルールにより得られた静止画像の推移に基づく時系列パターンを用いて動画像の検索を行なうことができる。例えば、野球中継であれば、“先行のチームのバッターがヒットを打ったシーン”などのようなコンテンツベースの検索が可能である。

3 クラスタリングによる連続した静止画像からの時系列データ生成

動画像は静止画像の時系列データであると考え、動画像を連続した静止画像に変換し、その静止画像のカラー情報を用いてクラスタリングを行なう。これにより各静止画像の特徴付けを行ない、静止画像の推移をクラスタの特徴で表現する。そこで、各静止画像の特徴の流れで現される時系列データに変換する。

3.1 静止画像のクラスタリング

本研究で使用したデータはプロ野球中継のムービーで、画像サイズは 360×240 、時間は 100 分間である。これから抽出する静止画像は、画像サイズはそのままの 360×240 、フレームレートは 1 秒間に 1 枚であり、静止画像は 6206 枚となる。

このような膨大な静止画像に対してそれらが持つ特徴等のコンテンツ情報を用いて検索を行なう場合に、連続した静止画像の特徴として、

1. 選手の位置や動作といったように頻繁に変化を伴う画像データである。
2. 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、このように頻繁に形状が変化し、位置の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、Kohonen の自己組織化マップ (Self-Organizing Map : 以下 SOM とする) [4] により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って発見、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。

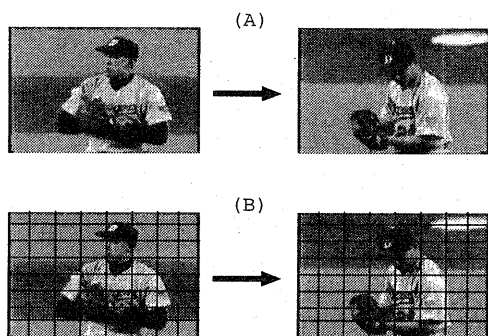


図 2: クラスタリングの問題点

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 2 の (A) に示すように同じ場面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 3 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れる可能性が高くなると考えられる。ただし、この手法にも問題点はある。今回使用した野球中継のような色数が少なく画面構成が単純なものではうまくいくが、色数が多く画面構成が複雑なものを対象とした実験では、余り良い結果が得られなかった。

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 (12×18)、1 回目の SOM によって特徴付けさ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。これによりブロックの出現頻度に基づいたクラスタリングを行なう。クラスタリングを行なう手順をまとめるといかなようになる (図 3 参照)。

stage1 パターンセルのクラスタリング

- step1 すべての画像を $N \times M$ のブロックに分割する。
- step2 ブロックのピクセルを SOM の入力ベクトルとして学習を行ない、SOM マップの生成を行なう。
- step3 元画像を SOM マップクラスタのアドレスで表現する。この特徴ブロックをパターンセルと呼ぶ。

stage2 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、学習を行なう。

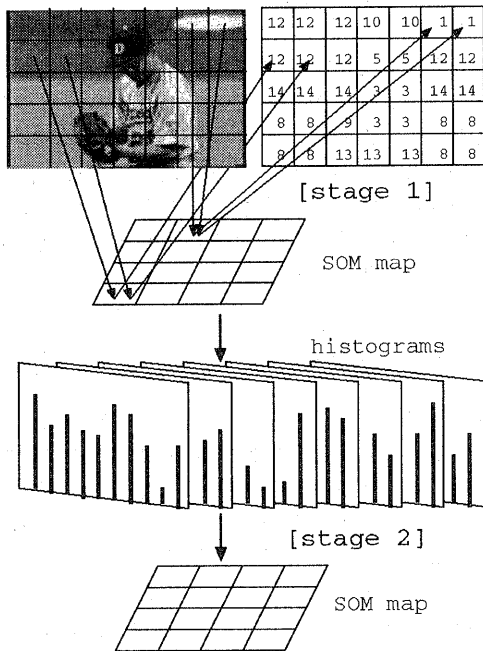


図 3: SOM による静止画像のクラスタリング

本研究では、プロ野球中継約 100 分間を実験対象とした。静止画像は 1 秒間に 1 枚を取り出し、約 6200 枚の連続した静止画像を実験データとした。画像のサイズは (360 × 240)、ブロック一つのサイズは (20 × 20)、ブロック数 N × M は、216 (18 × 12) とし、SOM マップは、1 段階目、2 段階目ともに、5 × 5 としてクラスタリングを行なった。この分類結果から得られた画像データを見てみると、類似した特徴 (同じ場面) の静止画像がクラスタリングされていた。本手法では物体の正確な位置情報を扱っているわけではないので、右投げ投手と左投げ投手などを区別することはできないが、今回の実験ではまずまずの結果が得られている。

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を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。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、

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

と表される。クラスタリングを行なった静止画像 6206 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 972 枚、クラスタに入った適合画像は合計 5161 枚、クラスタに入った不適合画像は合計 1045 枚であった。この結果を表 1 に示す。この分類の精度が時系列データからのパターン抽出を大きく左右する。

適合率	再現率
83.2%(5161/6206)	84.2%(5161/6113)

そして、それぞれのクラスタに見られる場面を視覚的に判断し、クラスタの特徴付けを行なった。この過程において、同じクラスタにクラスタリングされた画像に対して同じ気象の特徴を与えるという方法を適用した。先にも述べたが、クラスタリング方法の性質上大まかな分類しかできないため、クラスタの特徴付けも大まかにしかできないが、大まかな特徴付けはうまくいっている。この結果を表 2 に示す。この例では各クラスタの SOM マップアドレスを記号 (A~Y) で表現している。

クラスタ	場面
A,B,F	内野のプレー
C	バッターが打席へ向かうところ
D	電光掲示板
E	選手のアップ
I,T,W,Y	マウンドからの視点
J,K,N	打席に入った巨人のバッター
L	中日のピッチャーのアップ
M	打席に入った中日のバッター
O	バッターが打席をはずしているところ
Q,X	ピッチャーのアップ
S	中日ベンチ
V	外野のプレー

3.3 時系列データの生成

クラスタリングを行ない、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動的な分類を行なうことが可能になる。ここで、

連続した静止画像に割り当てられた標本となるクラスタの値を時間軸上に順に列挙する。これを図4に示す。

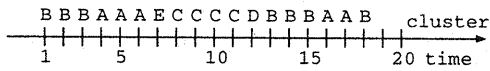


図4: クラスタの移り変わり

ここで、同じ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規則性や関係を導き出すことにより、大量の動画像データの中に存在するパターンを導出する。

4 エピソードルールの抽出

エピソードルール [5] は、連続的なデータ集合を解釈するために使用される。本稿では、静止画像の時系列データである動画像データにおいて局所的な相関ルールを選び出すために共出現対抽出アルゴリズム [8] を用いて、このエピソードルールの抽出を行う。

4.1 エピソードルール

時系列データからのデータマイニングアルゴリズムであるエピソードルールに関して簡単に述べる。エピソードルールは時系列データの中に起こるイベントのシーケンスを解析し、ある特定の時間に起こるイベントの組をエピソードとする。

図5に示すように、シーケンスにおいて連続的に起こるイベントの長いストリングが与えられた場合に、(E,31) (F,34) (A,35) (B,37) (C,38), ..., (D,49) といったように表現される。ここでA,B,Cは任意のイベントであり、それぞれにそのイベントが起こった時間が与えられている。簡単な例として、図5のようにAの後にBが起こるといったものやEとFの後にCが起こるといったように接近して起こるイベントの組み合わせをエピソードとする。このとき、どれだけ接近してイベントが起こるかを定めるためにウィンドウと呼ばれる間隔を導入し、これをシフトさせることによりある間隔の中に接近して起こるイベントの組み合わせを取り出すことによってエピソードの候補を抽出す

る。そしてある頻度以上のものをエピソードルールとする。

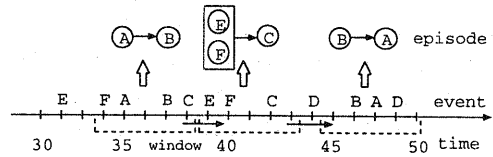


図5: イベントのシーケンスとエピソード

4.2 動画像におけるエピソードルール抽出

静止画像の時系列データである動画像データに対して、あるイベントAの後にあるイベントBが起こるといったシーン(意味的なまとまり)となるエピソードルールを抽出する手法について述べる。クラスタの推移からなる動画像の時系列データが与えられたとき、連続して出現するクラスタを同じ活動パターンを持つ連続した画像とみなして、それをシーケンスにおけるイベントとする。このときイベントは開始時刻と終了時刻をもつ時間区間であり、 $e_i = \langle C, S_f, T_s, T_e \rangle$ ($i=1, \dots, n$) と表現される。ここでCは連続して出現するクラスタ、 S_f はその連続度、 T_s, T_e はイベントの開始および終了時刻である。またこれらのイベントによって構成されるシーケンスをSとすると $S = (e_1, e_2, \dots, e_n)$ となる。図6におけるA', B', C'は連続度 S_f を2以上とした場合のイベントのシーケンスを表したものである。

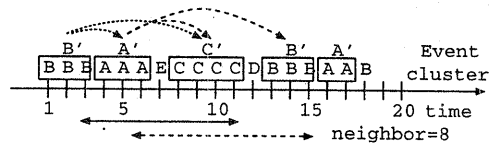


図6: 静止画像の推移におけるエピソードルール

そして、 B', A', C', B', A' というイベントのシーケンスから順に $\langle B'-A', B'-C' \rangle$, $\langle A'-C', A'-B' \rangle$, $\langle C'-B', C'-A' \rangle$, ... といったようにある間隔で近接して出現するイベントの対を抽出していく。ここで、イベントの対を抽出する間隔を定めるために近傍距離(時間的な近さ)を導入する。このときエピソードEは、 $E = \{[e_i, e_j], neighbor\}$ ($i=1, \dots, n-1, j=1, \dots, n$) と表現される。ここで $[e_i, e_j]$ は順序関

係 $i < j$ を満たす2つのイベントの組み合わせであり、*neighbor* は近傍距離である。この近傍距離はエピソードルールのウィンドウと類似した概念であるが、あるイベント A の後にどれだけ近接して次のイベント B が出現するかを定めるための時間区間であり、そのウィンドウの中に存在するイベントの組み合わせをすべて求めるといった操作は行わない。また、我々の手法では、エピソードの各々望ましい形をあらかじめ与えておいて時系列データからそのエピソードを探すのではなく、局所的な時間区間に近接して頻繁に共出現するイベントの組み合わせ集合を分類するために共出現対の手法を用いる。そのため次の共出現関係の式を適用し、イベントのシーケンスから抽出された候補となるエピソードに結合度を与える。

$$cohesion(e_i, e_j) = \frac{E_f(e_i, e_j)}{\sqrt{[f(e_i), f(e_j)]}}$$

ここで $f(e_i)$, $f(e_j)$ は各イベントの頻度、 $E_f(e_i, e_j)$ は共出現の頻度を表す。そしてこの *cohesion* に閾値を与えて、それ以上のものをエピソードルールとして取り出す。この *cohesion* は、ある2つの属性の共出現する関係の強度を与えるものであり、この例では2つのイベントの出現度が高くて共出現の頻度が低ければ、*cohesion* の値は小さくなり、逆に2つのイベントの出現度がそれほど高くなくとも、共出現の頻度が高ければ *cohesion* の値が高くなり、共出現する関係が高いといえる。

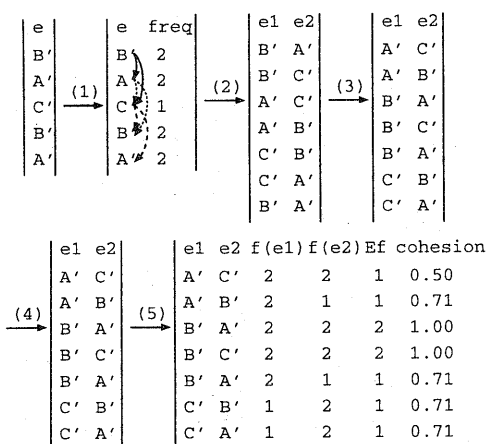


図 7: エピソードの抽出過程

図 7 に、クラスタの推移におけるエピソードの

抽出過程を示す。 e, e_1, e_2 はイベント、*freq* はイベントの頻度、 $f(e_1), f(e_2)$ はそれぞれ e_1, e_2 の頻度を表す。また E_f はエピソードの頻度を表す。各数値は図 7 の数値に対応している。

1. 各イベントの出現頻度 *freq* を求める。
2. 近傍距離の閾値を定めてイベントの組み合わせを求め、エピソードの候補を抽出する。図 6 の例では近傍距離を 8 以下とした。
3. } 辞書式順序でソートする。
4. }
5. 候補となるエピソードの頻度と結合度を求める。

以上の過程により、イベントの並びの中でより関係の強いエピソードを抽出する。(近傍距離内ではなく、あるイベントの個数までイベントの組み合わせを求め、エピソードの候補を抽出する方法も考えられる。その方法は現在進行中である。)

このように連続したクラスタをシーケンスにおけるイベントとし、同じクラスタに分類された画像を同じイベントである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シーンとなるエピソードを抽出する。これによりあるイベント A の後にあるイベント B が見られるといったように前後関係を反映したシーンを抽出する。また、エピソードの候補をとった後、近傍距離内にさらにイベントがある場合はそのイベントも加えて、3つのイベントからなるエピソードの候補の抽出も行った。

今回の実験では、連続度を 2 以上、近傍距離を 10 から 50 まで、また *cohesion* の閾値を 0.1 以上として、エピソードルールの抽出を行った。2つのイベントからなるエピソードルールの抽出結果を表 3 に示す。

表 3: 近傍距離とエピソード (*cohesion* > 0.1)

近傍距離の幅	10	20	30	40	50
エピソード数	1906	3880	5886	7880	9813

実験で対象とした野球中継の動画像データは 1 秒 1 フレームの静止画像の時系列データであるため近傍距離の幅 30 では、30 秒以内に現れるイベントを抽出していることになる。

考察

近傍距離の幅が短かすぎると満足にエピソードが抽出できず、また長すぎると重複したエピソードが増えることが考えられる。実験結果から今回の実験データでは、近傍距離の幅 30 で野球中継で試合を構成する主なイベントを抽出でき、エピソードルールにより試合中で起こる一般的なシーンの抽出を行うことができた。また、3つのイベントからなるエピソードの抽出結果でも同じ結論を得た。

5 シーン検索

得られたエピソードルールをデータベースに格納し、Servlet, JDBC, PostgreSQLを組み合わせ、Servletを使ったデータベース連携による順序関係を反映したシーン検索を行う Web アプリケーションの構築を行った¹。本システムは、ビデオの内容に基づいて多種多様なシーンを検索可能である。

5.1 Web アプリケーションの構築

まず、得られたエピソードを構成するイベントの値、開始値と終了値となる画像の ID, *cohesion* をデータベースに格納する。そして、図 8 に示すような Web アプリケーションを構築した。DougaForm.html から SQLResult.html まで Servlet, JDBC を介してデータベースへアクセスするところを時系列データからの時系列検索、SQLResult.html から DougaView.html までを検索結果の動画像表示となる。Web アプリケーションは、2つのイベントからなるエピソードと3つのイベントからなるエピソードからシーン検索を行えるものを用意した。

5.2 検索例

図 9 にシーン検索を行うシステムの検索画面を示す。このシステムではインニングの表裏と前後のイベントを選択することにより検索を行う。ここでイベントはクラスタリングによって分類された結果の特徴付けを行ったものを検索条件としている。図の上方の入力フォームでインニングの表裏とイベン

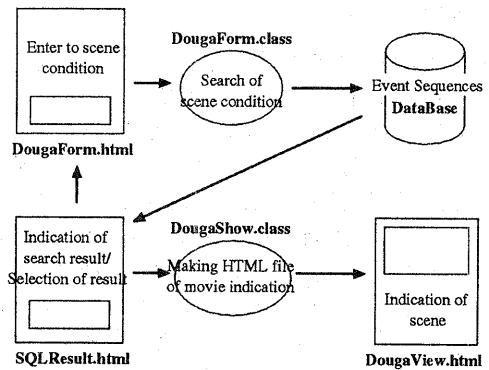


図 8: Web アプリケーション

トを組み合わせ検索することにより、ピッチャーが投げるシーン、ヒットを打ったシーン、バッターが空振りしたシーンなどが検索可能である。この例では、“先行のチームのバッターがヒットを打ったシーン”の検索を行っている。この結果、図の左下に14個のシーンが検索結果として得られた。検索結果の数字は開始値と終了値となる画像のIDを表している。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。この例では、中日の門倉投手から巨人の清水選手が2塁打を打ったシーンが表示された。

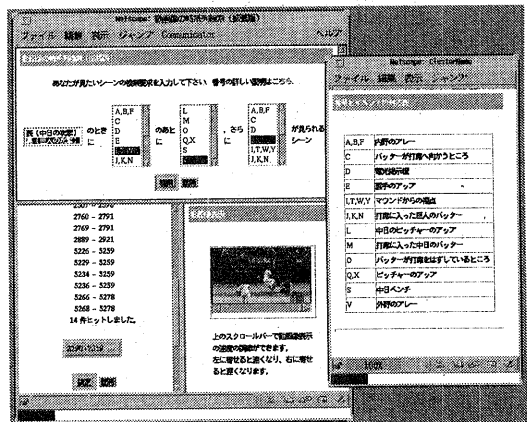


図 9: シーン検索の例

¹ <http://zeus.is.kochi-u.ac.jp/~takimoto/java/servlets/index3.html> で公開中

6 おわりに

本論文では、時間的連続性を持った時系列データである気象画像に対して有用であった、クラスタリング（自己組織化マップ）と相関ルール（エピソードルール）というデータマイニング法を組み合わせて、順序関係を反映した特徴パターンの抽出を行う手法を静止画像の時系列データである動画像（野球中継）に適用した。これにより抽出したパターン情報をデータベースとして蓄積し、シーン（意味的なまとまりの）を検索可能な動画像データベースの構築を行った。これによって野球中継で試合を構成する主なイベントを抽出でき、エピソードルールにより試合中で起こる一般的なシーンの抽出が行えることから、動画像でも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。

従来の手法 [7] では、パターン情報（イベント）を手を用いて対応づけているためデータベースの構築や維持にコストがかかるといった問題点があるが、本手法では特徴量を自動分類し、その分類結果に基づいて、コンテンツ検索を行うことができる方法を与えているため、人手をかけないことから膨大なデータにも適用でき、自動分類の精度が高ければ非常に効果的である。またエピソードの抽出により有用な情報を得ることができることから、情報圧縮の方面での適用も考えられる。

本論文では野球中継を例としたが、野球中継以外にも、色数が少なく構図がはっきりしている動画像に適用可能である。今後の課題として、

1. 画像全体を一つとしたクラスタリング方法をとるブロック DCT → SOM 方式を用いて作成する時系列データを加えた、2つの時系列データを用いてのルール抽出
2. Allenによって提案された時区間関係 [6] を利用した、より複雑なシーン検索への対応などがある。

参考文献

- [1] Fomg, J.(Edt.), "Data Mining, Data Warehousing & Client / Server Databases", Proc. 8th Int. Database Workshop, Springer (1997).
- [2] Alex, A.F., and Simon, H.L., "Mining Very Large Databases with Parallel Processing", Luwer Academic Publishers (1998).
- [3] R.Agrawal, T.Imelinski 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).
- [4] Kohonen. T., "Self - Organizing Maps", Springer, (1995).
- [5] H.Mannila, H.Tovinen and A.I.Verkanon, "Discovering frequent episodes in sequences", In First International Conference on Knowledge Discovery and Data Mining(KDD'95), 210-215, AAAI Press(1995).
- [6] J.Allen, "Maintaining knowledge about temporal intervals," Comm. of ACM, vol26, no.11, pp.832-843, 1983.
- [7] 牛尼剛聡, 広部一弥, 渡辺豊英, "利用者の視点に基づくシーン検索のためのイベント-アクティビティモデル", 電子情報通信学会論文誌, D-I, Vol.J82-D-I, No.1, pp.256-267, (1999)
- [8] 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, No2, pp179-189, (1989)
- [9] 片山幸治, 小西 修, "知識発見支援のための気象画像データベースの構築", 情報処理学会論文誌データベース, Vol.40, No.SIG 5(TOD 2), pp.69-78, (1999).
- [10] 片山幸治, 小西 修, "気象画像の時系列データ変換とエピソードルールの発見", DEWS'99, p.3C-3(p.8), (1999).