

気象衛星画像に対する時空間知識発見支援システムの構築 自己組織化マップによるインデクシングと時間依存ルールの発見

Construction of Spatial and Temporal Knowledge Discovery Support System for Weather Satellite Images

瀧本 裕一[1], # 本田 理恵[1], 小西 修[1]
Hirokazu Takimoto[1], # Rie Honda[2], Osamu Konishi[1]

[1] 高知大・理・数理情報

[1] Dept. Math. & Info. Sci, Kochi University, [2] Information Sci., Kochi Univ.

気象衛星ひまわりの日本付近の時系列可視画像 2 年分を対象に時間・空間知識の発見支援システムを検討した。画像は教師なし学習アルゴリズムの 1 つである自己組織化マップ (SOM) を用いてクラスタリングした。2 段階ブロック化 SOM の使用により、84.6-86.7% という高い精度のクラスタリングが実現できた。得られたクラスタ ID を展開した時系列データから連続したイベントや時間ウィンドウ内で大きな頻度で発生するイベント対の相関ルールを検出した。画像や得られた知識をデータベースに格納して SQL 検索インターフェースと検索結果の動画表示機能を実装することにより、ユーザーのインタラクティブな知識発見を支援できるシステムを構築した。

近年、大規模な科学データから知識を発見するシステムの要求が高まっている。我々は、気象衛星ひまわり (GMS-5) の 2 年分の時系列可視画像集合を対象として、半自動的な時空間知識発見とインタラクティブな高次の知識発見を支援するシステムについて検討を行った。

今回扱うひまわりの可視画像で顕著な特徴は雲の分布パターンであり、特に日本付近の画像では、季節によって台風や前線などの典型的なパターンが存在することが先見的にわかっている。また、こうしたパターンが時間的にある一定の頻度で繰り返す可能性が高いことも容易に予想される。このような背景を考慮に入れ、画像集合から時間空間知識の発見は以下のプロセスから構成することを考慮した。

- (1) 自己組織化マップによる画像のクラスタリング。
- (2) 専門家によるクラスタの意味付け (季節と雲の特徴による)。
- (3) 画像系列をクラスタ ID の時系列として展開。
- (4) イベント、時間ウィンドウ内で大きな頻度で発生するイベント対 (ルール) の検出。
- (5) 抽出された知識データベースへの格納。
- (6) エンドユーザーの検索によるインタラクティブな知識発見。

ここで使用した自己組織化マップは 2 層のニューラルネットワークから成る教師なしの学習アルゴリズムである。入力層は多次元ベクトルの集合からなり、競合層は 2 次元のマップ状のユニットから構成される。競合層のユニットは、入力ベクトルと同じ次元の参照ベクトルを持つ。学習は、入力ベクトルと競合層の各ユニットの参照ベクトルの距離を計算して最短距離を持つユニット (勝者ユニット) を探索し、その近傍ユニットの参照ベクトルを入力ベクトルに近づくように変更することによって行われる。この結果、先見的知識無しに、多次元情報の距離に基づく類似関係を、2 次元の競合層ユニットの分布として学習することができるため、知識発見支援の手法として有用性が高い。またクラスタリングにも使用することができる。

実験には 1997 年、1998 年各 1 年分の画像を 8 時間ごとにサンプリングしたデータセット (各 1044 枚、966 枚) を使用した。各画像は学習の入力ベクトルとしては、画像の各ピクセルの輝度値をラスタスキャン上に並べたものを入力ベクトルとして使用すればよいが、移動する前線・低気圧などの特徴を位置によらず同じタイプの画像としてグループ化させなければならないという問題がある。そこで各画像をブロック化し、各ブロックをデータ単位としてクラスタリングした後、画像内に出現したクラスタ ID のヒストグラムを再度クラスタリングするという 2 段階の学習方法を構成した。

実験では原画像を 12×18 のブロックに分割し各段階で 4×4 のマップを用いて 10000 回程度の学習を行った。学習後、各クラスタに分類された画像を参照し、エキスパート的な視点で季節、雲の特徴で意味付けしたところ、冬型の西高東低、夏の太平洋高気圧など一般的に知られるパターンときわめて類似したクラスタリングが行われている事がわかった。また、クラスタリングの精度 (適合率) は 84.6-86.7% という高い値を達成することができた。

得られたクラスタ ID を時系列データとして展開した後、連続して生起するクラスタ ID 列をイベントとして抽出し、さらに可変の時間ウィンドウ内で頻繁に発生する 1 対のイベントシーケンスを抽出した。時間ウィンドウを 80 時間 (3 日) から 400 時間 (ほぼ 2 週間) まで変化させたところ抽出されたルールは 10 個前後から 200 個程度までに及んだ。

こうして得られた知識 (クラスタ系列やルール) を画像と共にデータベースに格納した。ここでは効率的な

探索を可能にするため、時間（季節、月）に関する包含関係に基づき R-tree によるインデクス付けを行った。さらにユーザーに対して単純な条件や SQL によって検索が行えるインターフェースと検索結果の画像系列の動画表示機能を実装した。これにより、たとえば以下のような検索を行って結果を動画表示することが可能になった。

- ・ 1997 年 7 月 16 日から 20 日以内に発生した台風を検索せよ。
- ・ 台風発生後に起こる気象変化（ルール）を検索せよ。
- ・ 前線と台風が 5 日間発達した後、10 日以内に台風が発生する気象変化を検索せよ。

このような機能の実装により、データ集合からの半自動的な知識発見とともに、それを踏まえたユーザーによるインタラクティブな知識発見の枠組みを実現することができたといえる。なお本システムのデモについては以下の URL で参照することができる。

<http://zeus.is.kochi-u.ac.jp/~takimoto/java/servlets/index2.html>