

隠れマルコフモデルによる気象画像の時空間変動パターンモデリング

Spatial-temporal pattern modeling of a weather imagery by Hidden Markov Model

本田 理恵[1]; 勝吉 進一[2]; 小西 修[3]

Rie Honda[1]; Shinichi Katsuyoshi[2]; Osamu Konishi[3]

[1] 高知大・理・数理情報; [2] 高知大・理・数理情報科学; [3] はこだて未来大・複雑系

[1] Information Sci., Kochi Univ.; [2] Math. and Inform. Sci., Kochi Univ.; [3] Dept. Complex Systems, Future Univ. Hakodate

<http://www.is.kochi-u.ac.jp/~honda>

地球科学ではリモートセンシングやシミュレーションなどにおいて膨大な時空間データを得ることが多い。このような時空間データの評価は、従来時系列画像の可視化や、専門家の先見的知識に基づく各時間の特徴量の時系列データの評価により行われてきた。しかしこのような方法はデータの量が膨大になると非効率的になり、また新しい重要な特徴量を見逃すといった問題も考えられる。時空間データから時間に関する統計的性質を利用して半自動的な要約、規則性の発見をすることができれば、時空間データの特徴と量を生かした新たな知識の発見を促進することが期待できる。本研究では気象衛星ひまわり (GMS5) が撮像した日本上空の自己組織化マップによって時系列符号に展開して隠れマルコフモデルで表現することにより、季節内、季節感変動などのパターンを抽出することを試みた。

自己組織化マップ (SOM) は教師無し学習の 1 種であり、2 次元のマップ上に特徴量の近似度を反映した投影を反復的に行いマップにデータ集合の特性を学習させると同時に、その結果をデータ集合のクラスタリング (グループ化) に用いる事ができる。本研究では自己組織化マップにより画像系列を時系列の符号 (クラスタ ID) に変換し、さらにこの符号系列を隠れマルコフモデル (HMM) によって表現した。隠れマルコフモデルは、マルコフ過程に従って遷移する“状態” (各状態の出現確率および状態間の遷移確率で定義される。非観測) と、各状態で生成される記号 (観測) の出現確率分布から構成されるモデルである。時系列気象画像に隠れマルコフモデルを適用する場合、状態は類似した画像をほぼ同じ確率で生成する期間を意味し、具体的にはある一定の気象期間 (梅雨、季節など) に相当するものとみなすことができる。

実験では、1997 年 - 2001 年の GMS-5 の近赤外 IR1 の雲画像 (高知大学・東京大学・気象庁提供) を 24 時間おきにサンプリングして用いた。画像系列は画像内のオブジェクト (この場合雲塊) が移動してもできるようにブロック化とヒストグラム化を用いた 2 段階の自己組織化マップによって行った (Honda et al. 2001)。画像系列は trial and error により 6x6 のマップを使用した計 36 個のグループに分けられた。ここで展開された画像の記号系列に対し隠れマルコフモデルを適用した。なお、状態数は未知であるため、2-10 の状態についてそれぞれモデルを求めベイズの情報量基準によって最適な状態数とモデルを選択した。

その結果、状態数 5 において最適なモデルが得られ、画像やその時間分布の評価から状態と季節の対応を求めると (春、秋 真冬 春、秋、梅雨 夏、秋 盛夏) という、従来の認識とは異なる季節の組み合わせとしての状態とその遷移系列として表現された。隠れマルコフモデル自体はある状態がほかの任意の状態に遷移できることを仮定しているが、得られたモデルはとなりあった状態の間でしかほとんど遷移がおこらない鎖状の遷移系列であった。

このことは、画像特徴とその時間的発生系列の特徴だけに注目すると従来 1 つとみなされた季節にも複数の状態が存在し、また、特にそれらが複数の季節にわたって発生することを示している。これらは特に日本の気象の場合、春、夏 (盛夏を除く) 秋に顕著であったが、盛夏、冬に関しては従来の季節と一致したような状態としてモデル化された。また、得られた各状態における各記号の確率分布を自己組織化マップ上で可視化することにより、時空間の両方を考慮に入れたクラスタリングの可視化を行うこともできた。

なお、このモデルは確率分布をもつグラフで構成されているため、予測への応用の可能性もある。今後は、高次特徴の検出、予測への応用、他の地球科学の観測、シミュレーションデータに対しても検討していきたい。