

時空間データからの相関イベントクラスタの共起性の抽出

森 啓太[†] 本田 理恵[†]

[†] 高知大学 〒780-8520 高知県高知市曙町 2-5-1

E-mail: [†] {b113k302,honda}@is.kochi-u.ac.jp

あらまし 近年様々な分野で時系列画像などの時空間データが大量に蓄積されている. このような大量の時空間データから, ある地点で特徴的な変動が起きた時に時空間上の近傍で類似した変動が起きることを予測することが出来れば, 様々な分野で活用することが可能であろう. 本研究では, 時空間データに対してある基準時系列を与えた時, 高い正または負の相関係数を示す時系列を抽出してクラスタ化し, さらにそのようなクラスタが共起する部分時空間領域を相関ルールの評価指標を用いて定量的に評価し, ホットスポットとして可視化する手法について検討した. また, この手法の有効性について, 気象画像への適用例を示しながら報告する.

キーワード 時空間データ, 相関係数, 相関ルール, 支持度, 確信度.

1. はじめに

地球観測衛星の画像やセキュリティカメラの画像など, 様々な画像が時間と共に大量に蓄積されている. このような時系列画像には様々な時空間変動パターンが含まれているため, 時間, 空間の様々な断面で特徴を抽出し, 時空間変動に関する知識を取り出すことが期待される. 例えば, 西前, 本田(2014)^[1]は Hadoop を用いて, 時系列画像から効率的に時系列を抽出して基準時系列に対する相関係数の計算を分散処理によって高速に実施する手法を検討し, 気象画像に対する適用例を報告している.

相関係数を手掛かりとして, ある地点で特徴的な変動が起きた時, 時空間の近傍で類似した変動が起きることが予測できれば様々な分野で活用できるだろう. 時系列画像を時空間の領域に分割し, 領域ごとの共起性を評価することが出来れば時系列空間での現象の共起性の時空間変動を定量的にとらえることが可能になると考えられる.

本研究では時系列画像中のある基準時系列と高い正または負の相関係数を示す時系列を持つ領域を抽出してクラスタ化し, そのようなクラスタが多く共起する領域を相関ルールの評価指標を用いて定量的に評価し, 可視化することを試みた.

2. 手法

図 1 に時系列画像データでの時空間相関分析の概念図を示す. 時系列画像中である時間 t を起点として時間ウィンドウ ω における画像の階調値の時間変動を調査し, 特徴的な基準時系列 $V_s = \{I(x, y, t + i) | i = 0, 1, \dots, \omega - 1\}$ が得られたものとする. この基準時系列と相関性の高い変動が起きている位置ずれ, 時間ずれのある座標 $(x', y') = (x + dx, y + dy)$, 時間 $t' = t + dt$ の参照点 $V_r = \{I(x', y', t' + i) | i = 0, 1, \dots, \omega - 1\}$ を求める. これらの相関係数の高い時系列同士の共起を時空間の領域ごと

に評価するものとする.

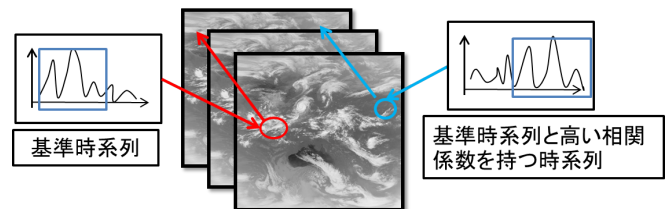


図 1. 時系列画像からの時空間相関分析の概念図

2.1. 相関分析

相関性の指標には相関係数 r を用い, 以下の式で求める.

$$r = \frac{\sum_{i=0}^{\omega-1} (V_s(i) - \bar{V}_s)(V_r(i) - \bar{V}_r)}{\sqrt{\sum_{i=0}^{\omega-1} (V_s(i) - \bar{V}_s)^2} \sqrt{\sum_{i=0}^{\omega-1} (V_r(i) - \bar{V}_r)^2}} \quad (1)$$

ここで $\bar{V}_s, \bar{V}_r$ はそれぞれ V_s, V_r の平均値とする. 時系列データにおいて相関係数は正の値だけでなく負の値も同様に重要であるため, 相関係数の絶対値が大きいものに注目して, 基準時系列と高い正または負の相関係数を持つ部分時系列を抽出する^[1]. 抽出された個々の部分時系列を イベント と呼ぶことにする.

2.2. イベントクラスタの抽出

隣接するイベントは同一の現象と考えられるので, 三次元画像のラベリング処理^[2]によって時空間内で塊状に存在するイベントをまとめてクラスタ化し, イベントクラスタ とする.

2.3. 共起性の時空間分布の評価

図 2 に共起性の時空間分布の評価方法の概要をまとめる. イベントクラスタの共起性は時空間の領域毎に異なると考えられるので, 時空間を小領域(ブロック)

に分割し、それぞれのブロック内でのイベントクラスタの共起性を調査する．さらにイベントクラスタの共起を走査する窓(時間，空間方向)を定義し，この窓をトランザクションとしてブロック内で操作させることでイベントクラスタの単発発生件数と共起件数を求める．この時，イベントクラスタは広がりを持つので，その最小包含矩形の 25%が窓内に含まれた時にカウントするものとする．走査窓によって求めたトランザクションをもとに相関ルールにおける支持度と確信度を求めることで共起性を評価する．

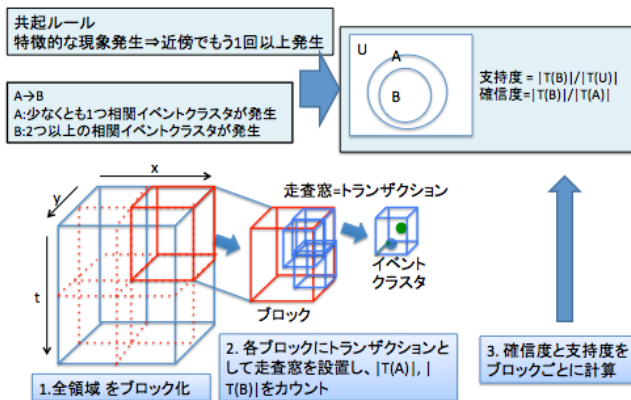


図 2. 共起性の時空間分布の評価方法の概要

相関ルールは事象の共起のしやすさを表すルールで，「 $A \rightarrow B$ 」という相関ルールは， A という事象が発生すると B という事象も発生しやすいことを意味する^[3]．このとき， A を含むトランザクションを $T(A)$ ， B を含むトランザクションを $T(B)$ とすると，支持度と確信度は以下の式で表すことができる．

$$\text{支持度} = \frac{|T(A \cdot B)|}{|T(U)|} \quad (2)$$

$$\text{確信度} = \frac{|T(A \cdot B)|}{|T(A)|} \quad (3)$$

支持度は全てのトランザクション数に対する A と B 両方を含むトランザクション数の割合で，データ全体での共起の起きやすさをしめす．確信度は条件部となる A を含むトランザクション数に対する A と B 両方を含むトランザクション数の割合で，条件となる事象が起きた時の共起の起きやすさを示し，条件と結果のつながりの強さを表す．

イベントクラスタの共起性を示す相関ルールを考える場合，イベントクラスタが走査窓内で一つだけ発生する事象を A とし，イベントクラスタが走査窓内で複数発生する事象を B とすると，相関ルールの支持度，確信度は以下の式に簡単化して表すことができる．

$$\text{支持度} = \frac{|T(B)|}{|T(U)|} \quad (4)$$

$$\text{確信度} = \frac{|T(B)|}{|T(A)|} \quad (5)$$

この場合，支持度は全トランザクション数に対する複数のイベントクラスタが発生したトランザクション数の割合で領域内での共起の起こりやすさを表す．確信度は最低一つのイベントクラスタが発生したトランザクション数に対し複数のイベントクラスタ数の割合でイベントクラスタ同士のつながりの強さを表す．

これら支持度と確信度によって時空間の領域ごとの定量的に表された共起性を可視化するものとする．

3. 気象衛星画像での実験

3.1. 対象データ

支持度と確信度による評価手法の実データへの適用例として，時系列気象衛星画像における共起性の評価と可視化を行う．

実験には運輸多目的衛星ひまわり MTSAT-1,2 (ひまわり 6,7) が一時間ごとに撮影した 560×560 pixel の画像^[4]を 2012 年 9 月から 12 月までの 2912 枚使用する．画像は北緯 70° から南緯 70° ，東経 70° から西経 150° の範囲を図 2 のようにマッピングしている．この画像の 5×5 pixel の輝度の平均値を特徴量として $112 \times 112 \times 2914$ グリッドのデータに対して相関分析を行う．使用した時系列画像の一部の画像を図 3 に示す．

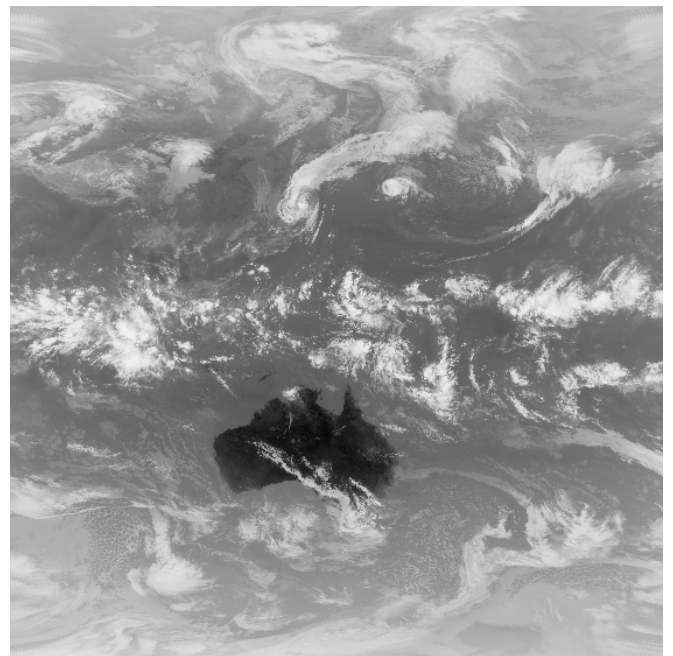


図 3. MTSAT-1 が撮影した 2012 年 10 月 18 日 3 時の画像

表 1 使用した画像の諸元.

項目	値
画像の種類	MTSAT-1,2 IR1
画像サイズ	560×560pixel
画像の範囲	北緯 70°～南緯 70° 東経 70°～西経 150°
画像枚数	2914 枚(2012.9~2012.12)
空間解像度	0.25°
サンプリング間隔	1 時間
基準時系列	西経 154.5°,南緯 7.0° 2012/9/28 から 10 日間

基準時系列は西経 176.25°，南緯 7.5°，画像上で (425,250)の点の 2012/10/10 から 10 日間のデータを用いた．基準点として選択した地点とその地点での雲の様子を図 4 に示す．基準点は熱帯収束帯という領域に位置する．熱帯収束帯とは赤道付近で南北両阪急からの貿易風が収束する帯状の領域であり，東西にのびた帯状の低気圧が発生し，その中で周期的に雲海が発生・消滅する．熱帯収束帯は太陽の位置に合わせて移動し，7 月から 1 月にかけて南下していく．

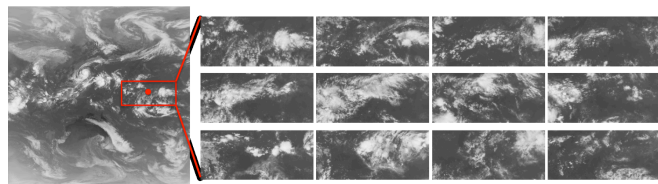


図 4. 2012 年 10 月 10 日 0 時(GMT)MTSAT-2 画像と赤い矩形領域内の画像の 20 時間おきの時間変化．赤点が基準点^[1]．

図 5 に基準点での時系列変動を示す．図 5 の縦軸は画像の階調値を表し，横軸が時間を表す．図 4 と図 5 から基準点付近ではほぼ 5 日おきに雲が帯状に発生・消滅を繰り返していることがわかる．よってこのような 5 日程度に 2 サイクル増減を繰り返かえすような変化が 1 度発生すると時間空間の近傍で共起する，といった領域と発生時期を検出しようとすることになる．

イベントクラスタ抽出のための相関係数絶対値の閾値は 0.7 とした．共起性を調査するための走査に関するパラメータは表 2 に示す．対象とするブロックサイズは空間方向に 28×28 グリッド，時間方向に約 1 ヶ月に相当する 720 グリッドとする．走査窓は空間方向に 14×14 グリッド，時間方向に約 10 日間に相当する 240 グリッドとした．ここで，ブロック，走査窓はいずれもそのサイズは時系列の始点に対して設定され

ているので，この場合，10 日間の時間遅れまで”時間的近傍”とみなして拡張された共起関係を探査することになる．走査の単位（間隔）は空間方向に 7 グリッド，時間方向に 1 日に相当する 24 グリッドとした．なお時系列画像からの時系列データの抽出に関しては Hadoop によるブロックごとの分散化による処理を用いた^[1]．

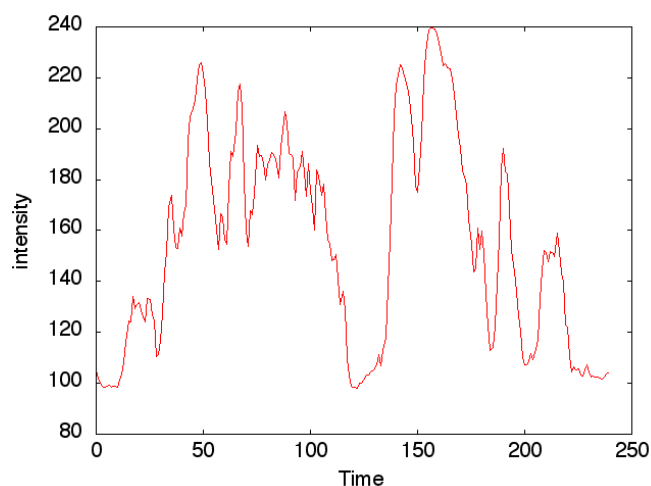


図 5. 使用した基準時系列．

表 2. 実験条件. サイズの単位はいずれもグリッド.

項目	値
グリッド数	空間方向 112×112 (5x5 pixel 毎に平均) 時間方向 2912
ブロックサイズ	空間方向 28×28, 時間方向 720(約 1 ヶ月)
走査窓 (トランザクション) サイズ	空間方向 14×14 時間方向 240(約 1 ヶ月)
基準時系列	時間方向 240

3.2. 実験結果

時空間の全領域から抽出したイベントクラスタの可視化結果を真上から見たものを図 6 に，斜め上から見たものを図 7 に示す．図の縦軸は画像の y 軸を横軸は画像の x 軸を表す．色はイベントクラスタが発生した月を表し，赤が 9 月，緑が 10 月，青が 11 月，紫が 12 月である．図 6 からある時間，空間に偏ってイベントクラスタが存在し，弧状に広がる大きなクラスタも存在することが分かる．

各月における領域ごとの支持度と確信度を図 8 に示す．9 月の結果を見ると，左から 3 番目，上から 2 番目のブロックで支持度，確信度がともに高いことが確認できる．これはイベントクラスタの共起が発生しやすく，一つのイベントクラスタが発生した場合 10 日以内に付近でイベントクラスタが発生しやすいこと

を示す．10月にはその右下のブロックに支持度と確信度の高いブロックが確認できる．よって共起が起こりやすい領域が時期によって移動したと考えられる．このように時空間の領域ごとにイベントクラスターの共起を評価しているのので，類似した変動の共起するホットスポットの時空間変化を定量的に，また視覚的に容易に確認することができる．

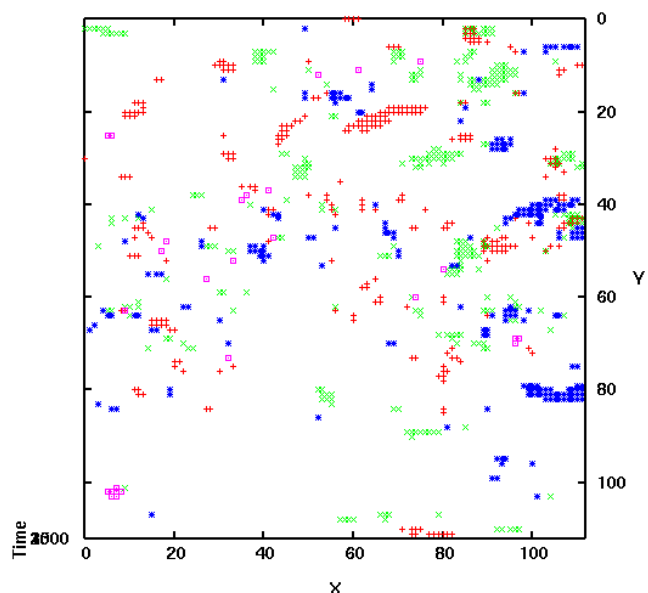


図 6. 月ごとのイベントクラスター(赤:9月,緑:10月,青:11月,紫:12月)

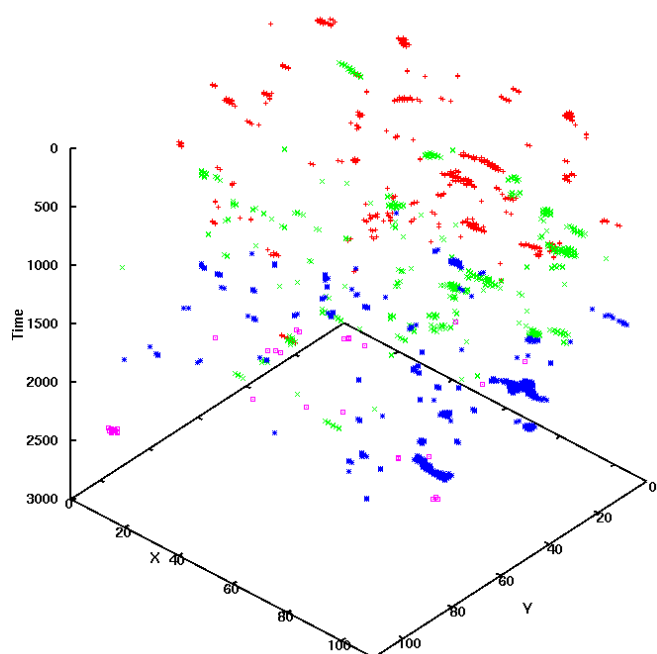


図 7. 月ごとのイベントクラスター(赤:9月,緑:10月,青:11月,紫:12月)

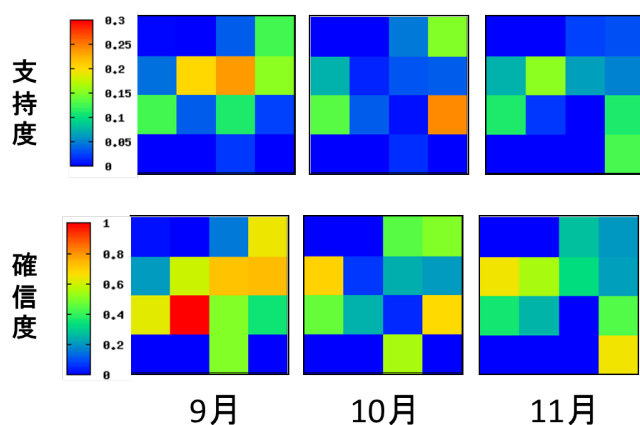


図 8. 2012 年 9～11 月の月ごとの支持度と確信度．支持度の値の範囲は 0～0.3，確信度は 0～1．

ここでターゲットにした熱帯収束帯(ITCZ)の例を図 9 に示す．熱帯収束帯の特徴は下記の通りであった．

- 南北両半球からの貿易風が収束する帯状の領域．周期的に雲が発生．
- 季節によってその位置が移動し，7 月から 1 月にかけては南下．

図 8 に示された可視化の結果は，赤道付近の領域で支持度と確信度が共に高い領域が存在し，また 9 月から 11 月 にむかって南下していく傾向をとらえており，その共起の特徴（支持度，確信度）を定量的にとらえていることから，熱帯収束帯における雲の周期的発生現象とその時期による移動をとらえ，かつ定量的に表現できていると考えられる．

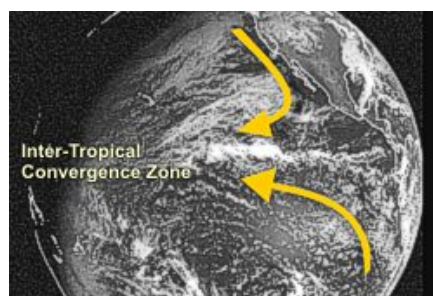


図 9 熱帯収束帯の雲形状^[5]．

4. まとめ

時系列画像でのイベントの共起性を定量的に評価し，可視化した．支持度と確信度の二つの評価指標を用いて領域内でのイベントクラスターの発生の特徴とその時間変化を知ることが出来た．この評価方法は気象画像以外の時系列画像や，時空間データへの応用が期待できる．

なお，今回は基準時系列を既知としたが，一般的には

基準時系列自体も未知であることが多い。このような場合、どのようにして特徴ある基準時系列を発見するかという問題については網羅的なクラスタリングや、分類などによって発見することを検討する中である。また、この手法を用いて共起の時空間変動の評価をする場合、扱う問題によってブロックの大きさや走査窓の大きさを変えて評価を行う必要がある。これらのチューニングを容易にする手法として、走査窓を自動的に伸縮させながらの評価などを今後の課題として検討していく。また、様々な大きさの走査窓に対する計算を行う際の計算量の増大が考えられるため、分散処理による効率的な計算の Hadoop, MapReduce による実装を検討していく。

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
その他の方々に感謝する。

参 考 文 献

- [1] 西前ほか, Hadoop による時系列画像分散データマイニングシステムの検討, DEIM Forum 2014, D1-6, 2014.
- [2] MIST Project, <http://mist.murase.m.is.nagoya-u.ac.jp/trac/>, 2010.
- [3] R. Agrawal et al., Mining association rules between sets of items in large databases, Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93, 207-217, 1993
- [4] 高知大学気象情報頁,
<http://weather.is.kochi-u.ac.jp>.
- [5] Inter-Tropical Convergence Zone, National weather service, <http://www.srh.noaa.gov/jetstream//tropics/itcz.htm>, 2010.