

令和 2 年 6 月 16 日現在

機関番号：34315

研究種目：基盤研究(C) (一般)

研究期間：2016～2019

課題番号：16K00414

研究課題名(和文)細胞の状態を表現する確率モデルの構築と細胞核の動態解析への応用

研究課題名(英文) Development of stochastic models to represent cellular states and their application to analysis of nuclear division dynamics

研究代表者

遠里 由佳子 (Tohsato, Yukako)

立命館大学・情報理工学部・教授

研究者番号：80346171

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：線虫の野生胚やRNAi胚の大規模な2次元タイムラプス顕微鏡画像データを画像処理して得た細胞核分裂動態の時空間定量データを解析に用いる(野生胚：N=1, RNAi胚：対象遺伝子549種類, 1遺伝子あたり約N=3)。時空間定量データから特徴抽出し、得られる時系列データからHMM(Hidden Markov Model)などの確率的生成モデルの構築を試みた。そして、時系列のパターン認識と分節化、野生胚の時系列を正常データとして学習したモデルを用いた異常検知の実現を目指した。

研究成果の学術的意義や社会的意義

時系列データの分節化は、音声データやモーションキャプチャーデータなどでは重要な課題となっている。線虫の初期発生の時系列データから、HMMを基盤とするさまざまな確率的生成モデルを構築することで、時系列の分節化、認識、異常検知を実現し、人工データやシミュレーションデータを用いて、それらの分節化の精度比較を行なった。生命科学の分野で産出される非線形・多変量の時系列データに対して、こうした時系列の分節化や認識、異常検知のアプローチは、新たな計算表現型解析の実現を意味している。

研究成果の概要(英文)：We focused on spatiotemporal quantitative data of cell division dynamics obtained by image processing from two-dimensional time-lapse microscopy movies of *Caenorhabditis elegans* embryos with large-scale (N=1 wild-type embryo, about N=3 RNAi embryos for each of 549 target genes). We tried to construct probabilistic generative models such as HMM (Hidden Markov Models) from time-series data by extracting features from such quantitative data. By using constructed models, we aimed to realize pattern recognition and segmentation of time series, and anomaly detection using the models in which time-series data of wild-type embryo were learned as normal data.

研究分野：バイオインフォマティクス

キーワード：隠れマルコフモデル バイズ推定 確率的生成モデル 時系列データ解析 表現型解析 *C. elegans* データ駆動型解析 システム生物学

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

生命科学の分野では、大規模・精密な撮影を可能にする顕微鏡計測技術の開発と、畳み込みニューラルネットワークなどを導入した高精度な画像処理技術の発展により、物体検出や追跡で得られた膨大な核や細胞といった生命現象の「時空間情報を数値として含む定量データ」が得られつつある。こうした定量データから特徴抽出することによって得られる多変量時系列データは、生命現象を動的システムとして認識し解析する上で欠かすことができない。しかし一方で、決定論的な力学系の変化から非線形な性質を持つ。時間変化にはゆらぎがみられる。そうした性質が解析を困難にしている。そうした問題に対し、センサによって得られたモーションキャプチャデータや音声データに対するアプローチが解決法になりえると考えた。そして、多変量時系列データに状態遷移を仮定し物理的時間に基づきモデル化する、隠れマルコフモデル(HMM: Hidden Markov Model)に着目した。なお、研究開始当初、生命科学の分野では1分子計測で得られるデータに対しHMMの利用例が報告されているのみであった。

2. 研究の目的

細胞の動態を顕微鏡計測し画像処理することで得られる「時空間情報を数値として含む定量データ」は、生命現象を動的システムとして認識し解析する上で重要な高次元情報である。しかし、さまざまな画像処理アルゴリズムの開発が進む一方で、得られた時空間定量データの解析法の検討は未だ限られている。そこで、データ駆動型サイエンスの実現にむけた新しい方法論の開発を目指して、本研究課題では線虫の初期発生の時空間定量データ(図1左)から細胞核の特徴を抽出し、特徴抽出で得た多変量時系列データ(図1右)を用いて、ベイズ統計による確率生成モデルの構築に挑戦した。

研究の目的は、得られた多変量時系列データに対する、パラメータに依存しない頑強な時系列パターンの認識や正確な時系列の分節化、異常検知を実現すること、であった。そして提案手法を、線虫野生胚や1つの遺伝子をノックダウンした胚の初期発生の多様な細胞核動態データに適用することにより、一時的な時間変化に着目した表現型解析、細胞核動態の制御機構に欠失した遺伝子が担う機能の推定を可能にすることを目指した。

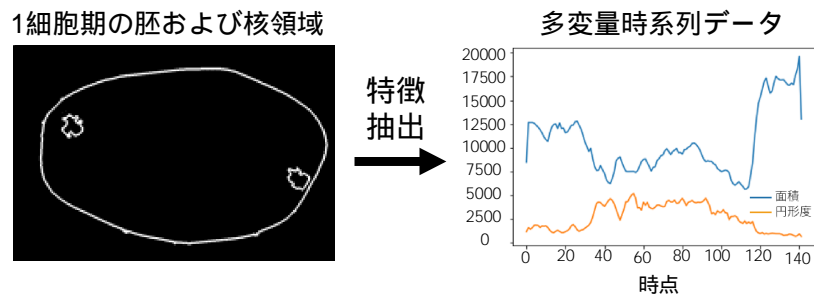


図1: 線虫の1細胞期の細胞核の時空間動態

3. 研究の方法

線虫の初期発生の野生胚や、対象とする遺伝子の発現をRNAi (RNA interference; RNA干渉)で抑制したRNAi胚(野生胚: N=1 定量データ, RNAi胚: RNAi対象遺伝子549種類, 1遺伝子あたり約N=3 定量データ)から特徴抽出して得られる多変量時系列データを解析の対象とした。それらデータに対し、尤度が最大になるよう確率的生成モデルのパラメータを学習することで、(a)時系列パターンの認識、(b)時系列の分節化、(c)正常な時系列データを学習した確率的生成モデルを用いた異常検知の実現を目指した。確率的生成モデルとして、HMMや、HMMを状態遷移に時間継続長を考慮するよう拡張したHSMM(Hidden Semi Markov Model)、それらにディリクレ過程(HDP; Hierarchical Dirichlet Process)を追加することにより状態数の自動推定を行うHDP-HMMやHDP-HSMMを研究の対象とした。

4. 研究成果

細胞核から核の大きさや円形度を抽出した多変量時系列データ(図1)に対して、細胞周期に相当する状態パターンの学習と時系列の分節化を実現するため、最初に、野生胚の時系列データを入力として状態数を指定しHMMを構築し、AIC(Akaike information criterion)で、最適な状態数を判定するアプローチを試みた(図2,3)。その結果、欠損値の位置の関係なく、分節化が行われることを確認した。さらに、細胞周期は一般に、M期(M phase)と、間期(interphase)に分けられ、間期はさらに、G1期、S期、G2期に分けられるが、それらの細胞周期に相当する状態割り当てが行われていることを確認した。

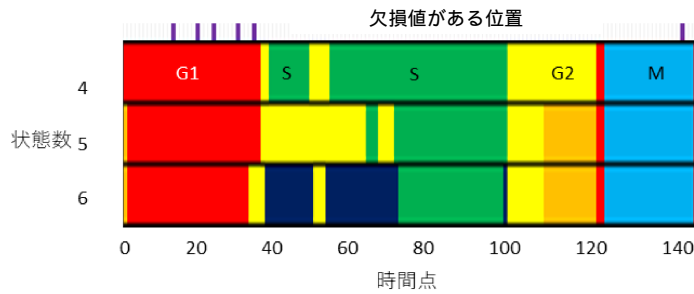


図 2：状態数を 4 から 6 に変化させたときの Ergodic 型 HMM の時系列に対する状態の割り当て結果。色が推定された状態を表しており、各状態に対応する細胞周期 G1, S, G2, M 期が示されている。

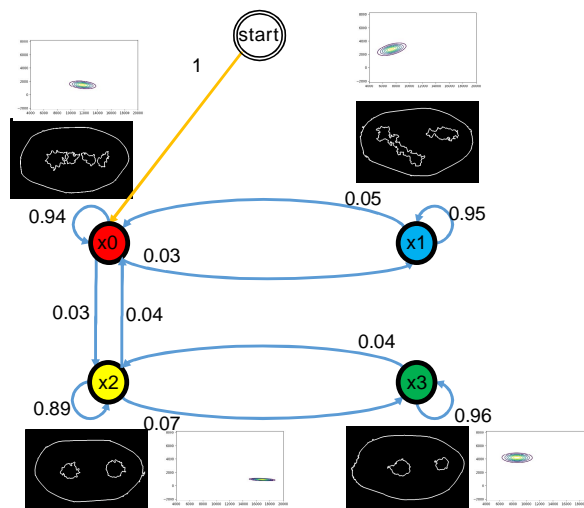


図 3：図 2 で構築した HMM のうち状態数 4 の HMM の詳細。x0 から x3 の状態に割り当てられた色は、図 2 の図の色と対応している。各状態には、代表的な細胞状態と、出力分布(2 次元正規分布)が示されている。

しかし、RNAi 胚データに対してモデルを構築する場合には、細胞周期などに多様な変化があることが考えられるため、状態数を指定することは望ましくない。そこで、状態数も自動で推定できる HDP の導入を検討した。加えて、HMM には、状態から別の状態に費やされた時間を考慮しないため、細胞の状態を推定する上では不正確な推定となつてきている可能性があった。そこで、滞在時間分布をモデル化できる HSMM を導入とした。結果として、HDP-HSMM を実現した。

次に、分節化の精度を評価するため、正解が明らかなデータとして、(1)指定した分布からランダムサンプリングして作成した短い時系列を繋ぎ合わせた人工データおよび、(1)先の実データで構築したさまざまな確率的生成モデルから生成データを作成した。そして、人工データや生成データを HDP-HSMM や HDP-HMM に入力として与えることにより、HDP による状態数の自動推定の精度、HSMM の効果の評価を行なった(図 3)。そして、Left-to-right 型 HMM から生成した時系列を入力として HDP-HSMM は HDP-HMM と比べて高い精度の分節化が得られ(HDP-HSMM: 98.6%、HDP-HMM: 85.2%) HSMM の効果により、より自然な状態変化が推定できることを確認した。

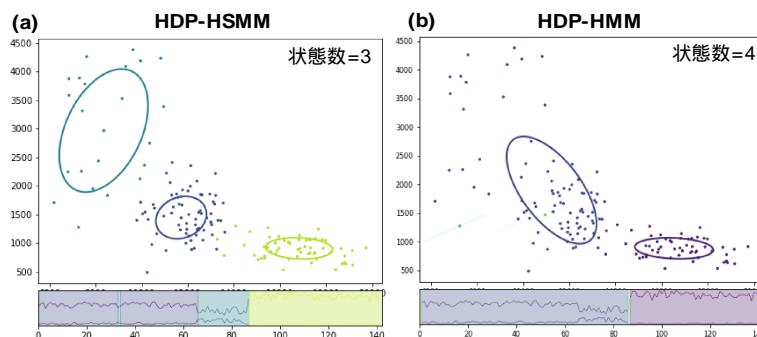


図 3：生成データに対する HDP-HSMM(a)と HDP-HMM(b)の状態割り当て例

さらに、細胞核のラベル情報を用いて、1細胞期からの4細胞期ではまでの細胞分裂を、分岐のある時系列として抽出しHMMに導入する方法を提案した。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Yang Sihai, Xianhua Han, Yukako Tohsato, Koji Kyoda, Shuichi Onami, Ikuko Nishikawa, Yenwei Chen	4. 巻 24
2. 論文標題 Phenotype analysis method for identification of gene functions involved in asymmetric division of <i>Caenorhabditis elegans</i>	5. 発行年 2017年
3. 雑誌名 Journal of Computational Biology	6. 最初と最後の頁 436-446
掲載論文のDOI（デジタルオブジェクト識別子） 10.1089/cmb.2016.0210	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Xian-Hua Han, Yukako Tohsato, Koji Kyoda, Shuichi Onami, Ikuko Nishikawa, Yen-Wei Chen	4. 巻 8
2. 論文標題 Nuclear detection in 4D microscope images of developing embryo using enhanced probability map of top-ranked intensity-ordered descriptors	5. 発行年 2016年
3. 雑誌名 IPSJ Transactions on Computer Vision and Application	6. 最初と最後の頁 8
掲載論文のDOI（デジタルオブジェクト識別子） 10.1186/s41074-016-0010-3	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 Yukako Tohsato, Hatsumi Okada, Koji. Kyoda, Shuichi Onami
2. 発表標題 Data-driven analysis of female pronuclear migration by image-processing embryonic movies in <i>Phenobank</i>
3. 学会等名 22nd International C. elegans Conference (国際学会)
4. 発表年 2019年

1. 発表者名 遠里 由佳子, 岡田 初美, 京田 耕司, 大浪 修一
2. 発表標題 公共データベースの顕微鏡画像を用いた細胞核分裂の定量計測と表現型解析
3. 学会等名 生命医薬情報学連合大会 (IIBMP2018)
4. 発表年 2018年

1. 発表者名 遠里 由佳子, 岡田 初美, 高山 順, 京田 耕司, 大浪 修一
2. 発表標題 公共データベースの顕微鏡画像を用いた核分裂動態の定量計測とデータ駆動型解析
3. 学会等名 バイオイメージ・インフォマティクスワークショップ2016
4. 発表年 2016年

1. 発表者名 Yukako Tohsato, Hatsumi Okada, Jun Takayama, Koji Kyoda, Shuichi Onami
2. 発表標題 Developing quantitative resource for computational analysis from images of C. elegans embryogenesis in a public database Phenobank
3. 学会等名 The Allied Genetics Conference (TAGC) 2016 (国際学会)
4. 発表年 2016年

〔図書〕 計1件

1. 著者名 京田 耕司, ホー ケネス, 糸賀 裕弥, 遠里 由佳子, 大浪 修一	4. 発行年 2020年
2. 出版社 可視化情報学会	5. 総ページ数 4
3. 書名 生命現象の時空間動態情報の可視化	

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考