

プロジェクト名： イメージデータ解析

プロジェクトディレクター： 松井 知子 教授（統計数理研究所）

[1] 研究プロジェクト

(1) 目的・目標

本プロジェクトでは、最先端科学分野における複雑な現象を視覚的に表し、そのイメージデータを統計的手法により深く高度に解析して、様々な意思決定支援に役立てる基盤技術を確立することを目的とする。最先端科学分野としては、医用や生体、社会科学をはじめ、「生命、地球、環境、人間・社会」に関する幅広い分野を対象とする。複雑な現象を視覚的に表して人間の直感を働かせやすくすることにより、その現象の本質に迫ることを狙う。

現象の視覚表現は、データコーディングまで含め、分野固有の知識やスキルを利用して行う必要があるために、技術要素の多くは分野依存であることが想定される。ただし、異なる分野の視覚表現技術は互いに参考になると考えられる。一方、視覚的に表したイメージデータの高度解析技術は分野間で共有でき、分野横断的に有用できることが想定される。本事業では、最先端科学の幅広い分野での展開を通じて、複雑な現象の視覚表現およびそのデータコーディング技術のノウハウを蓄積し、汎用的なイメージデータの高度解析に関わる方法論を創出することを目指す。

具体的に、医療・生物関連の「バイオイメージング・インフォーマティクス」、「オプティカルイメージングデータを用いた脳の動作原理の解明」のサブテーマ1と2、人間・社会関連の「人間・社会の視覚情報データ解析」のサブテーマ3を取り組む。各サブテーマの目標を次に示す。

— サブテーマ1 「バイオイメージング・インフォーマティクス」の目標

ベイズ統計学や機械学習を方法論の基軸にバイオイメージデータ解析の様々な問題に取り組む。細胞や分子レベルのバイオイメージング・テクノロジーが急速に発展したことで、生物学や医学の諸関連分野において、静止画や動画像データの解析手法に対するニーズが高まることとなった。これにより近年、当該分野では機械学習やコンピュータビジョン、統計科学の先進技術が急速に導入されている。このような学術分野は、生物情報学のコミュニティでは、バイオイメージ・インフォマティクスと呼ばれている。当該分野の典型的なデータの特徴として、多数のオブジェクト（細胞や粒子）が高密度に分布し、位置や形状が時空間的な変化を示すこと、さらに顕微鏡光学系に起因するノイズ、データの時空間解像度の低さ、撮影条件の違いや個体差によるオブジェクトの形状・サイズの不均一性が挙げられる。このような複雑性を持つ画像データの解析は、統計科学の問題としてチャレンジングである。本プロジェクトでは、ノイズ処理、特徴抽出、セグメンテーション、物体認識、形状解析、トラッキング、イメージアライメント（レジストレーション）等、バイオイメージデータ解析の幅広い課題に取り組み、バイオイメージ・インフォマティクスの基盤となる先端技術を創出し、また新たな応用分野を戦略的に開拓していくことをを目指す。

— サブテーマ2 「オプティカルイメージングデータを用いた脳の動作原理の解明」の目標

呼吸活動に関与するニューロン、抑制性ニューロン、アストロサイトの定量的、かつ客観的な検出アルゴリズムの開発、相関、因果性の解析方法の開発、ならびに、自発的同期活動時の細胞活性化順序の穏やかな秩序の解明を行う。

— サブテーマ3 「人間・社会の視覚情報データ解析」の目標

人間・社会の人や物の映像データから、イベント検出・予測や人や物のトラッキングなどを行うことにより、潜在的に重要な（動的）因子を発見する方法を確立することを目指す。本方法を汎用的なツールとして整備することにより、医療・生物関連データの解析等にも利用可能とすることを狙う。

(2) 必要性・重要性（緊急性）

医療・生物関連分野では、生物学や医学の研究開発プロセスにおいて、イメージデータの解析は最早必要不可欠なピースとなっている。顕微鏡観察に基づく細胞の網羅的フェノタイピング、蛍光イメージングによるタンパク質の発現量および時空間ダイナミクスの定量解析、1分子計測・分析など、イメージ解析技術の利用機会は枚挙に暇がない。しかしながら、実験技術としての革新・波及速度に比べて、データの解析技術およびそれを活用できるヒューマンリソースの供給が圧倒的に遅れている。これは国内外に共通する実情である。

また、ニューロンやアストロサイトの自励的同期現象は脳内の情報伝達において重要な脳機能であり、このメカニズムを探るための注目すべき同期現象の一つとして脳幹における周期的呼吸活動がある。最近では、カルシウムイメージング法を用いた多細胞同時記録など先端的な測定技術が開発され、得られる知見は神経生理学のみならずニューロインフォマティクスの分野でも重要な役割を担いつつあるが、イメージングデータにおいてニューロン、抑制性ニューロン、アストロサイトなどの客観的な検出方法は確立しおらず、細胞レベルでのネットワーク推定に関する研究例はほとんどなされていない。

人間・社会関連分野では、クラウドコンピューティングの普及が進む中、様々な現象の収録映像が利用可能となっている。これらの映像を深く解析し、新価値を見出すことにより、防災や減災、および豊かさを享受できる社会の実現に資することが期待されている。

(3) 期待される成果等（学問的効果、社会的効果、改善効果等）

複雑な現象の視覚表現およびそのデータコーディング技術のノウハウの蓄積、および汎用的なイメージデータの高度解析技術の確立が期待できる。

特にサブテーマ1に関しては、細胞や分子のイメージング技術が発展し、生物学や医学の研究で大量の画像データを解析する機会が増加し、当該分野において、機械学習や統計科学の先進技術が急速に導入されている。本研究の目的は、ベイズ統計および機械学習の方法論に基づいて、細胞や粒子のセグメンテーション、数え上げ、幾何学的特徴の抽出、形状解析、定量化を行うための新しい方法論を構築することである。研究成果の適用分野は、分子生物学の基礎研究から、合成バイオロジー、医療・健康分野における創薬・診断・治療など、多岐に渡る。

またサブテーマ2に関しては、正常な呼吸リズムの生成・維持は、生命の維持に必須な身体システムである。呼吸リズムは、下位脳幹部に存在する呼吸中枢内での神経細胞の自発的同期活動により生成されており、呼吸リズムに関連する細胞（呼吸細胞）の個々の挙動等については研究が進んでいるが、自発的同期活動が生成する神経ネットワーク機構自体については明らかになっていない。

(4) 独創性・新規性等

最先端科学分野における複雑な現象を視覚的に表したイメージデータとして扱う点、またイメージデータを高度解析する技術を分野横断的に研究開発する点が独創的である。さらに、イメージデータ解析では新規性の高い統計的手法を導入・開発していく。

特にサブテーマ1に関しては、現在の細胞・分子イメージングのデータ解析の多くは、実験系の研究者が独立に既存の画像処理ツールを用いて試行錯誤しながら問題に対処しているというのが現状である。本研究は、バイオイメージ・インフォマティクスの研究課題の内、特徴抽出、セグメンテーション、形状解析、トラッキング、定量、システム解析に目標を定め、統計学の方法論の礎を築き、同時に当該分野に対して積極的な技術還元を行う。バイオイメージングのデータの大きな特色は、複雑な形状を持つ大量のオブジェクトが画像空間上で高密度に分布し、さらに、それらが時間・空間的に変化する点である。このような複雑性を持つ動画像データの解析は、統計科学の問題としてもチャレンジングである。

またサブテーマ2に関しては、神経ネットワークに関する実験系の研究で、個々の細胞の挙動、もしくは最大でも数個程度の細胞の同時解析、または活動エリアの検出などのマクロレベルの解析が主に行われている。個々の細胞の活動がどのように連携してネットワーク全体の機能を生みだすのかについては、重要なテーマであるものの、これまでの知見は少ない。本研究では、統計科学的方法論研究と神経生理学の専門家がチームを構成し、神経生理学的には神経ネットワークシステムと細胞レベルの活動の関連について生理学的な指標を提供し、また多細胞同時解析により大規模データの相関関係から因果関係に迫るという新たな手法の確立を目指し、「種々の細胞の活性化順序ならびに細胞の特定機能の有無」と言う個々の細胞の活動レベルと「ネットワーク全体の情報伝達経路・順序の秩序形成」と言う全体の機構を結び付けようとする点が本研究の最も独創的な点である。

(5) これまでの取り組み内容の概要及び実績

サブテーマ1では、線虫神経系のカルシウムイオンイメージングのデータをテストケースとして、細胞体の位置検出、細胞数の自動計数法、セルトラッキングやイメージレジストレーションのアルゴリズム開発を推進してきた。ベイズ統計やスペースモデリング、最適化の先進技術がこれらの方法論面の基盤を担っている。一連の成果については、国内では2013年度統計関連学会連合大会において最優秀報告賞に選出され（報告者：徳永旭将、2015年3月まで統計数理研究所 特任助教、2015年4月より九州工業大学 准教授に着任）、国外では、生物情報学のトップカンファレンスの一つである ISMB2014 (The 22nd Annual International Conference on Intelligent Systems for Molecular Biology) の講演者に選出された。また、同じく生物情報学のトップジャーナルである Bioinformatics 誌に論文一報が掲載された。さらに、人材育成および啓蒙活動の一環として、2014年度統計数理研究所・公開講座「バイオイメージングデータの統計解析」の開講や当該分野に向けた講演活動を通じて、定量生物学やシステムズバイオロジーのコミュニティに向けて積極的に情報発信を行ってきた。

またサブテーマ2では、最近の神経計測技術の進歩により脳神経活動に対する様々なモダリティによる計測法が開発されている。そのなかでもカルシウムイメージング法は数ミリ秒から数十秒の周期で生じる神経集団の同期現象や興奮の伝播を多細胞同時蛍光ライブイメージングとして計測することを可能とした。しかし、データは複雑な多次元の構造を持ち（2次元の計測面、時間軸、計測条件、繰り返し測定）、また、カルシウム感受性色素の退色や還流液による振動などに起因するによる強いドリフト成分やアーチファクトが重畠しており、大規模解析においては解決すべき多くの要素がある。研究チームでは、異なる動物サンプルの脳幹断面画像を共通のプラットホームで解析するための標準画像の作成方法の開発、ピクセル間の計測時間差の補正方法、蛍光イメージングデータを用いたアストロサイトと吸息性ニューロンの識別方法の開発などを行ってきた。

さらにサブテーマ3では、TRECVID（映像索引付け・検索、イベント検出に関する国際的な競争型評価プロジェクト）に参加して上位の成績を示すなど、イメージデータ解析に関する高度な技術を有している。さらにガウス過程による音楽情報処理、無線センサーネットワーク解析を行うなど、時空間モデリングに関する高度な技術を有している。

(6) 国内外における関連分野の学術研究の動向

サブテーマ1に関しては、バイオイメージ・インフォマティクスの世界的な動きは非常に活発化している。学会誌の特集号 (Nature Methods, 2012; BMC Cell Biology)、総説の発表 (Bioinformatics, 2008)、Paper Submission Track の新設 (Bioinformatics, 2012)、国際会議の大規模化・多様化 (International Workshop on Bioimage Informatics; International Conference on Systems Biology (ICSB); Cold Spring Harbor Laboratory Meeting (CSHL)) など、当該分野の急速な発展を反映した動きは活発であ

る。これに呼応する形で近年日本において、バイオイメージングの研究ユニットが相次いで設立されている（理化学研究所生命システム研究センター、理化学研究所分子イメージング科学的研究センター、大阪大学免疫学フロンティア研究センターなど）。

サブテーマ2に関しては、呼吸リズムは、呼吸中枢内での同期した神経活動で生成され、哺乳類の延髓腹外側には parafacial respiratory (pFRG) と preBötzinger complex (preBötC) と呼ばれる呼吸リズム生成に関する二つのペースメーカー領域が存在する。この領域における呼吸リズム生成についての従来の研究では、mouse の vivo 実験による抑制性ニューロンの寄与、preBötC を含むスライス標本（呼吸スライス）による in vitro の系での吸息性のリズム活動の生成、呼吸スライスを用いた preBötC における興奮性ならびに抑制性ペースメーカーニューロンの発見などの報告がある。また、近年、複数電極を用いた preBötC での局所フィールド電位 (LFP) の記録とその発火パターンや発火頻度に基づいた解析から、呼吸リズム形成に関連するニューロン及びアストロサイト（以下、呼吸細胞という）が神経ネットワーク内で自発的同期活動を行う際には細胞の活性化が確率的に起こっており、そのネットワークダイナミクスは従来考えられていたより柔軟であるとの報告がなされた。しかしながら、これらは呼吸細胞の種類・数、単一呼吸細胞の挙動、ネットワークダイナミクスに注目した研究であり、自発的同期活動に必要な神経ネットワーク構造やネットワーク内の呼吸細胞の活性化順序などについてネットワーク全体を俯瞰しながら多数の呼吸細胞の活動を調査した研究はこれまでに行われていない。

[2] 研究計画

(1) 全体計画

最先端科学分野の複雑な現象を対象とする次の三つのサブテーマを取り上げる。具体的にそれらのサブテーマに取り組みながら、視覚表現およびそのデータコーディング技術のノウハウを蓄積し、汎用的なイメージデータの高度解析の方法論を確立する。各サブテーマで研究開発する技術の関係性について検討し、汎用的なイメージデータ解析の方法論の創出に向けた指針を得る。

各サブテーマの全体計画を示す。

– サブテーマ1 「バイオイメージング・インフォーマティクス」の全体計画

細胞・分子のイメージングデータにもとづくオブジェクト検出、オブジェクト・トラッキング、イメージアライメント等の新規アルゴリズムを開発する。さらに、画像情報にもとづく特徴抽出、分画、分類、可視化など、バイオイメージ・インフォマティクスの幅広い研究課題に取り組む。

– サブテーマ2 「オプティカルイメージングデータを用いた脳の動作原理の解明」の全体計画

実験手法に局所細胞外電位 (LFP) と 2 光子顕微鏡によるカルシウムイメージングを用いて、2 方針に基づいて行う。まず、局所細胞外電位 (LFP) で計測される呼吸活動とカルシウムイメージングデータの各ピクセルに対応する時系列との間で時間依存相互関解析を行い、呼吸活動に関与するニューロン、アストロサイトを検出し、自発的同期活動時の細胞の活性化順序の秩序形成において果たす役割について調べる。次に、検出されたニューロンにおいて呼吸活動を予測する時系列モデルを構築し、抑制性ニューロンとアストロサイトからの寄与を考慮した外生変数型時系列モデルへと拡張させる。そして、赤池情報量規準 (AIC) などの指標を用いて有意に動力学的な寄与を与える抑制性ニューロンやアストロサイトを検出し、因果関係にあるニューロン、抑制性ニューロン、アストロサイトのネットワーク構造を推定する。実験としてはレーザーアブレーション法によってペースメーカー細胞を神経ネットワークから除去して、その細胞活性化順序への影響について調べる。

– サブテーマ3 「人間・社会の視覚情報データ解析」の全体計画

人間・社会の人や物の映像データから、イベント検出・予測や人や物のトラッキングなどを行うことにより、潜在的に重要な（動的）因子を発見する方法を確立することを目指す。本方法を汎用的な

ツールとして整備することにより、医療・生物関連データの解析等にも利用可能とすることを狙う。

(2) 各年度の計画

各サブテーマでイメージデータ解析技術の研究開発に取り組む。さらに、各サブテーマで研究開発する技術の関係性について検討する。

平成 26 年度

- サブテーマ1：細胞検出・トラッキングアルゴリズムの機能強化ならびに高速化に取り組む。さらに、ライブセルイメージングの分野において新規ユーザーならびに適用対象を開拓し、開発手法の運用体制を整備する。
- サブテーマ2：局所細胞外電位（LFP）計測は吸息性時に同期して生じるニューロンの賦活（呼吸性バースト）を記録するためのものである。LFP の時系列を参照関数として用い、イメージデータの各ピクセル座標における時系列との相互相関解析において、同時相関、タイムラグを考慮した相関、呼吸バースト毎の相関の 3 種類の方法を考慮したアルゴリズムを開発する。神経生理学実験においては、自励的同期活動（吸息性活動）時のバーストタイミングの違いに基づく細胞種の分類を行うことを目標として研究を進める。
- サブテーマ3：本研究で扱う人や物の映像データセット、および特徴量の選定を行う。特徴量抽出を行い、データ整備を進める。同時に動的な状態空間モデル、およびその推定アルゴリズムについて、汎用的なツールとして整備していく。

平成 27 年度

- サブテーマ1：イメージアライメント・アルゴリズムの研究を重点的に推進する。
- サブテーマ2：アストロサイトの活動のオンセット、オフセットの検出方法や賦活波形の減衰率の推定方法を検討し、吸息性アストロサイトの検出方法を開発する。そして、吸息性ニューロンとアストロサイトとの相関性や因果性の推定を行い、ニューロン、アストロサイトの局所ネットワーク構造を明らかにする。局所ネットワーク構造の推定のためには、ニューロンの呼吸性バーストを予測する時系列モデルを構築し、抑制性ニューロンとアストロサイトからの寄与を考慮した外生変数型時系列モデルへと拡張させる。さらに、「抑制性ニューロン」の「自発的同期活動中における活性化のタイミング」ならびに「機能阻害が細胞活性化順序の秩序に及ぼす影響」についても解析を行う。
- サブテーマ3：本研究で扱う人や物の映像データセット、および特徴量の選定を行う。特徴量抽出を行い、データ整備を進める。同時に動的な状態空間モデル、およびその推定アルゴリズムについて、汎用的なツールとして整備していく。

[3] 研究推進・実施体制

- サブテーマ1 :

・研究代表者

〔統計数理研究所〕 吉田 亮

・共同研究者

〔統計数理研究所〕 逸見昌之

〔金沢大学〕 広瀬 修

〔九州工業大学〕 徳永旭将

— サブテーマ2 :

・研究代表者

[統計数理研究所] 三分一史和

・共同研究者

[自治医科大学] 菊地千一郎、西多昌規

[兵庫医科大学] 越久仁敬、尾家慶彦

[キール大学] Andreas Galka

[ゲッティンゲン大学] Swen Huelsman

— サブテーマ3 :

・研究代表者

[統計数理研究所] 松井知子

[国立情報学研究所] 佐藤真一

・共同研究者

[統計数理研究所] 玉森 聰

[国立情報学研究所] Duy-Dinh Le

[ロンドン大学] Gareth W. Peters

[4] 研究の進捗状況

— サブテーマ1 :

線虫 C. エレガンスの神経系は、302 個の神経細胞で構成され、神経回路網の配線図が全て明らかになっている。近年のカルシウムイオンイメージング技術の進歩により、全神経細胞の約半分に相当する約 150 個の神経細胞の活動状態を同時計測できるようになった。これにより、生きた線虫の神経ネットワークの情報処理の様子をリアルタイムに観察できるようになる。データは、空間 3 次元プラス時間軸の立体動画である。オブジェクトの形状（細胞核）は、細胞ごとに多少の違いはあるが、視覚的に違いを識別することは非常に難しい。また、頭部にとりわけ多数の神経細胞が密集しており、計測中に線虫が不規則に動くことで細胞の位置は時間変化する。したがって、データから各細胞の活動状態を推定するには、細胞数の数え上げ、位置検出、セグメンテーション、そして形状均一性が高い密集した 100～200 個の細胞集団を追跡するための画像解析技術が不可欠となる。本研究では、ベイズ統計学と最適化のテクニックにもとづき個々の問題に対する要素技術を開発し、パイプライン化を行った (ImageClimber02)。細胞の数え上げと位置検出のタスクでは、カーネル密度推定と Repulsive Parallel Hill-Climbing 法という独自の最適化アルゴリズムを組み合わせ、解析手法の設計を行った。細胞のトラッキングでは、マルコフ確率場で細胞位置の時空間変化をモデリングし、トッラカーの置換や併合を回避しながら、約 150 個の細胞の動きを追跡することに成功した。最後に、各時刻のトッラカーの位置情報から細胞を分画し、分画領域の輝度情報から活動状態の推定を行うことに成功した。

また、上記の手法 (ImageClimber02) の開発と並行して、粒子フィルタと呼ばれるオブジェクト・トラッキングの手法に基づき、もう一つ別の細胞追跡アルゴリズムを開発し、ソフトウェアの公開を行った (図 1) (SPF-CellTracker: <http://github.com/ohirose/spf>)。検証用データを多数用意して、ImageClimber02 と SPF-CellTracker の包括的な性能比較を行ったところ、現時点では SPF-CellTracker が高精度なトラッキングを達成することが判明した。次年度以降は、SPF-CellTracker を主力に切り替え、研究開発を推進していく予定である。

SPF-CellTracker

This is a software suite for tracking more than a hundred of cells from 4D live cell imaging data with following characteristics:

1. gray scale image,
2. globular-like cell shapes,
3. strong covariation of cells to be tracked,
4. no background image except background noise and salt- and-pepper noise.

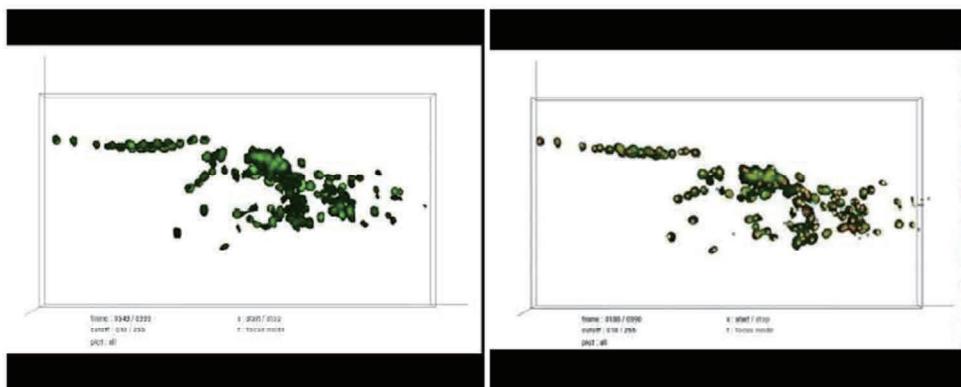


図1 細胞追跡ソフトウェア SPF-CellTracker のウェブサイト(<http://github.com/ohirose/spf>)

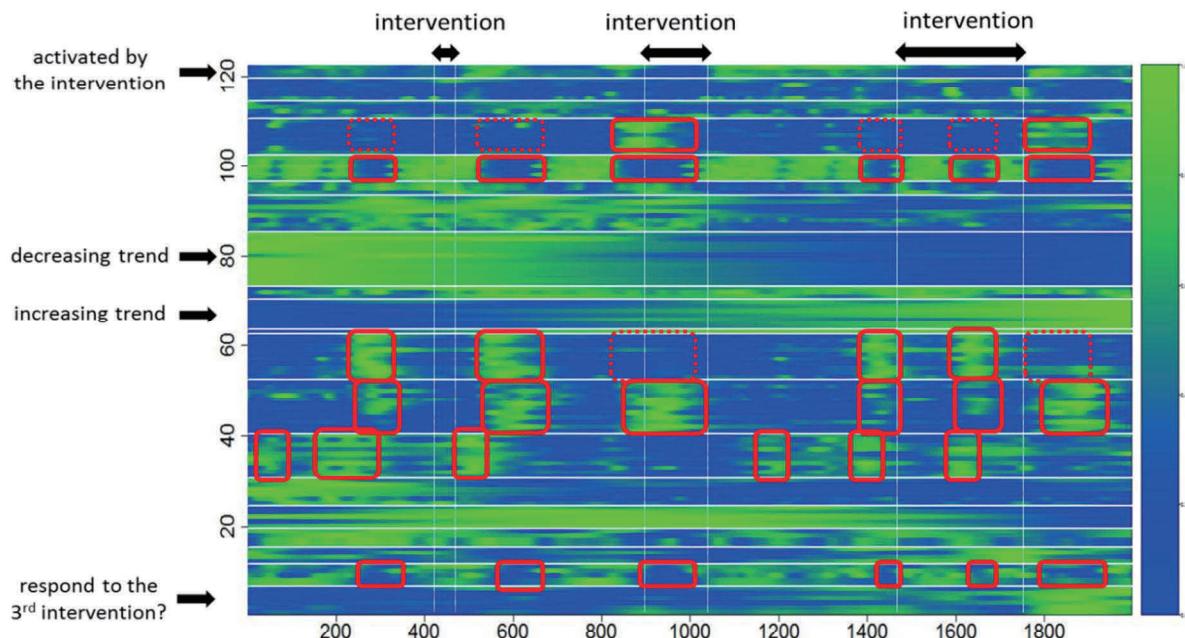


図2 4D イメージングデータの細胞トラッキングから得られた線虫の神経活動状態

現在、上記プログラムを用いて様々なデータに対して神経活動状態の定量化を行い、(a)刺激に応答するニューロン群の同定、(b)時系列情報に基づき神経系をネットワークレベルで理解するための、データ解析手法の開発を推進している（図2）。

さらに当該年度は、細胞アノテーションおよびイメージレジストレーションのアルゴリズム開発にも着手した。まず初めに、いくつかの既存の非剛体レジストレーションアルゴリズムを実装し、実データに基づく性能検証を行ったところ、これらはいずれも許容水準に程遠い性能であることが判明した。そこで、疎性学習に基づく細胞のマッチング手法やトポロジー保存を正則化に組み込んだ非剛体レジストレーションアルゴリズムの開発に着手することとなった。

– サブテーマ2 :

カルシウム指示薬である Oragon green BAPTA-1 (OGB-1) を利用した蛍光ライブイメージングデータの各ピクセルに対応する時系列の分散値のマップ、局所細胞外電位 (LFP) として計測される呼吸活動と蛍光ライブイメージングデータの各ピクセルに対応する時系列に時間依存相互相関解析を施して得られる相関マップ、そして、RFP (Red Fluorescent Protein) の強度マップにより、アストロサイトと吸息性のニューロンを区別する方法を開発した。

– サブテーマ3 : [添付資料1]

人間・社会の人や物の映像データから、イベント検出・予測や人や物のトラッキングなどを行うことにより、潜在的に重要な（動的）因子を発見するために、ガウス過程を用いた状態空間モデル、およびその推定アルゴリズムによるイメージデータ解析方法を開発し、汎用的な Monte Carlo Dynamic Classifier (MCDC) ツールとして整備を進めた。モーションキャプチャーのデータ (CMU Graphics Lab Motion Capture Database) を用いて、識別性能を評価する実験を行い、妥当な推定結果が得られることを確認した。

[5] 研究成果物

① 知見・成果物・知的財産権等

– サブテーマ3 :

1. Monte Carlo Dynamic Classifier ツールの改良版 [添付資料1]

② 成果発表等

<論文発表>

[学術論文]

– サブテーマ1 :

1. T. Tokunaga, O. Hirose, S. Kawaguchi, Y. Toyoshima, T. Teramoto, H. Ikebata, S. Kuge, T. Ishihara, Y. Iino, R. Yoshida, Automated detection and tracking of many cells by using 4D live-cell imaging data, Bioinformatics, 30(12):i43-51, 2014.

– サブテーマ2 :

2. Boiroux D, Oke Y, Miwakeichi F, Oku Y, Pixel timing correction in time-lapsed calcium imaging using point scanning microscopy, Journal Neuroscience Methods, 30,237,pp60-68, 2014.
3. Oku Y., Fresemann J., Miwakeichi F., Hülsmann S., Respiratory calcium fluctuations in low-frequency oscillating astrocytes in the pre-Bötzinger complex, Respiratory Physiology & Neurobiology 2015 (accepted).

4. Inoue H, Shimizu S, Nara H, Tsuruga T, Miwakeichi F, Hirai N, Kikuchi S, Watanabe E, Kato S, Basic Study for New Assistive Technology Based on Brain Activity during Car Driving@Journal of Robotics and Mechatronics@Journal of Robotics and Mechatronics,pp253-260, 2014.
— サブテーマ3 :
5. G. W. Peters, T. Matsui, F. Septier, and A. Tamamori, “Estimation and Calibration in Gaussian Process State Space Models: MCDC Tool”, Journal of Statistical Software (submitted).

[データベース] 特になし。

[著書等] 特になし。

[解説・総説] 特になし。

[その他] 特になし。

<会議発表等>

[招待講演]

— サブテーマ1 :

1. 徳永旭将 「バイオイメージ解析におけるベイズ統計の応用」、生命機能数理モデル検討会、大阪大学免疫学フロンティア研究センター、2014年5月28日
2. 吉田亮 「ライフサイエンス分野におけるベイズ統計の進展と応用」、データ科学の発展と植物科学との融合、東京大学中島薰一郎記念ホール、2014年11月28日
3. 寺本孝行、吉田亮「線虫中枢神経系のまるごと計測および定量解析に向けた取り組み」、定量生物学の会第七回年会、九州大学筑紫キャンパス、2015年1月11日
4. 吉田亮 「ライフサイエンス分野におけるベイズ統計の先端応用」、バイオスーパーコンピューティング研究会ウィンタースクール2015、休暇村伊良湖、2015年1月30日
5. 吉田亮 「ライフサイエンス分野におけるベイズ統計の先進応用」、第7回ミニ数理デザイン道場、東京大学大学院医学系研究科、2015年2月19日
— サブテーマ2 :
6. 三分一史和, 尾家慶彦, 光学的イメージング法とニューロン、アストロサイトの識別, 招待講演, 電子情報通信学会医用画像研究会”医用画像解析における統計数理的手法”, 2014.09.02.

[一般講演]

国際 :

— サブテーマ1 :

1. T. Teramoto*, T. Tokunaga, O. Hirose, Y. Toyoshima, Y. Iino, R. Yoshida, T. Ishihara, "Whole-brain imaging of C. elegans reveals multi-neuronal dynamics under non-stimulus condition", Wiring the Brain, New York, USA, 2015.3.
2. T. Teramoto*, Y. Toyoshima, T. Tokunaga, R. Yoshida, Y. Iino, T. Ishihara, "4-D imaging of neuronal activities in the whole central nervous system visualizes correlative patterns between multiple neuron", C.elegans Topic Meeting: Neuronal Development, Synaptic Function & Behavior, Madison, USA, 2014.7.
3. T. Tokunaga*, O. Hirose, S. Kawaguchi, Y. Toyoshima, T. Teramoto, H. Ikebata, S. Kuge, T. Ishihara, Y. Iino, R. Yoshida, "Automated detection and tracking of many cells by using 4D live-cell imaging data", 22nd Annual International Conference on Intelligent Systems for Molecular Biology

(ISMB2014), Boston, USA, 2014.7.

－ サブテーマ2 :

4. Okada Y, Fukusi I, Takeda K, Oyamada Y, Oku Y, Miwakeichi F, Someya H, Ishiguro M, Tamura Y, Pokorski M, Post-hypoxic potentiation of breathing is induced by sustained activation of astrocytes, The 54th Annual Meeting of the Japanese Respiratory Society, Osaka, Japan, 2014.04.26.
 5. Okada Y, Takeda K, Someya H, Oku Y, Miwakeichi F, Ishiguro M, Tamura Y, Pokorski M, Astrocytes mediate short-term plasticity of breathing, Experimental Biology 2015, Boston, U.S.A. 2014.3.28-2015.4.1.
- － サブテーマ3 :
6. T. Matsui, "Monte Carlo Dynamic Classifier (MCDC) Tool", STM2014, 2014.7.

国内 :

－ サブテーマ1 :

1. 広瀬修、川口翔太郎、徳永旭将、豊島有、寺本孝行、佐藤賢二、池端久貴、佐藤博文、久下小百合、石原健、飯野雄一、吉田亮 「3次元動画像内の非常に多数の細胞領域を自動追跡するための粒子フィルタ手法の開発」、人工知能学会全国大会（第28回）、愛媛県県民文化会館、2014年5月13日。
2. 徳永旭将、広瀬修、川口翔太郎、豊島有、寺本孝行、池端久貴、久下小百合、石原健、飯野雄一、吉田亮 「形状均一性が高い密集した細胞集団の位置検出およびトラッキング」、バイオイメージ・インフォマティクスワークショップ2014、岡崎コンファレンスセンター、2014年6月10日。
3. 池端久貴、吉田亮 「Repulsive Parallel MCMC アルゴリズムによる塩基配列のモチーフ探索」、2014年度統計関連学会連合大会、東京大学本郷キャンパス、2014年9月14日。
4. 広瀬修、徳永旭将、豊島有、寺本孝行、久下小百合、石原健、飯野雄一、吉田亮 「4Dライブセルイメージングデータ内の非常に多数の細胞を自動追跡するための空間粒子フィルタ手法の開発」、2014年度統計関連学会連合大会、東京大学本郷キャンパス、2014年9月14日。
5. 徳永旭将、広瀬修、吉田亮 「カーネル密度関数の局所空間変形にもとづくイメージアライメント」、2014年度統計関連学会連合大会、東京大学本郷キャンパス、2014年9月14日。
6. 池端久貴、吉田亮 「Repulsive parallel MCMC algorithm for discovering diverse motifs from large sequence datasets」、生命医薬情報学連合大会2014、仙台国際センター、2014年10月3日。
7. 徳永旭将、広瀬修、川口翔太郎、豊島有、寺本孝行、池端久貴、久下小百合、石原健、飯野雄一、吉田亮 「Automated detection and tracking of many cells by using 4D live-cell imaging data」、生命医薬情報学連合大会2014、仙台国際センター、2014年10月3日。

－ サブテーマ2 :

8. 三分一史和、尾家慶彦、統計値マップによるニューロン、アストロサイトの識別、統計数理研究所共同研究集会「ダイナミカルバイオインフォマティクスの展開 III」2014.09.18.

<受賞> 特になし。

③ その他の成果発表 特になし。

以上

ガウス過程状態空間モデルによる推定：MCDC ツールの紹介

Gareth W. Peters (UCL)

松井知子 (統数研)

Francois Septier (Telecom Lille)

玉森 聰 (統数研)

1. はじめに

時系列解析の分野で広く利用されるモデルに状態空間モデルがある。状態空間モデルは、ノイズを含む状態（遷移）方程式と観測方程式によって表される。両方程式の同時推定は、特にノイズが非線形ガウスの場合には難しい。ノイズが線形ガウスの状態空間モデル（線形ガウス状態空間モデル）では状態変数の推定にはカルマンフィルタが、ノイズが静的な場合には再帰的最小二乗推定法（勾配フィルタ）がしばしば用いられる。

一般に線形ガウス状態空間モデルは状態空間の非線形性を表現する能力は低い。加法的なガウスノイズの場合には拡張カルマンフィルタを適用することができ、また非ガウス性の加法的ノイズの場合にはunscentedカルマンフィルタを適用することができる。粒子フィルタや逐次モンテカルロ法などの手法が適用されることもある。先行研究の多くは状態方程式と観測方程式が既知であることを前提としているが、実際には両方程式は明らかでない場合もある。

我々はノンパラメトリックに定式化した状態空間モデルの新たなクラスについて検討した。このクラスでは、非線形なノンパラメトリックの状態空間モデルをフルベイズのアプローチで推定する。状態系列は未知、状態と観測変数に関する事前知識は十分になく、状態と観測関数は不確定である場合を想定している。状態と観測関数が次の状態空間上の滑らかな写像であることだけを仮定している。

$$f : \mathbb{R}^{\dim(\mathbf{x}_t)} \mapsto \mathbb{R}^{\dim(\mathbf{x}_t)}, \quad \text{and} \quad g : \mathbb{R}^{\dim(\mathbf{x}_t)} \mapsto \mathbb{R}^{\dim(\mathbf{y}_t)}.$$

この新しいクラスの状態空間モデルの特徴は、状態と観測関数がともにランダムな関数であり、それぞれが滑らかな連続関数の無限次元空間上のガウス過程により与えられることである。ガウス過程はノンパラメトリックベイズ法の一つである。ガウス過程は平均関数と（カーネル関数で表された）共分散関数により特徴づけられ、統計応用の分野で広く利用されている。共分散関数／カーネル関数についての研究は盛んに行われており、定常性、非定常性、周期性などの特徴をもつカーネル関数が提案されている。

本稿では、このガウス過程状態空間モデルに基づく統計モデリングを容易にするソフトウェア、Monte Carlo Dynamic Classifier (MCDC) ツールを開発したので紹介する。ガウス過程状態空間モデルは状態関数、観測関数のダイナミクスに事前分布を置くフルベイズの形で定式化される（次式）。これにより、複雑で動的な現象を捉えることができる柔軟なモデルのクラスが得られる。

$$\underbrace{\mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1}, \mathbf{e}_t) + \mathbf{n}_t}_{\text{Latent State Equation}} \quad \text{with} \quad f(\cdot) \sim GP(m_f(\cdot; \boldsymbol{\Phi}_f), k_f(\cdot, \cdot; \boldsymbol{\Theta}_f)),$$

$$\underbrace{\mathbf{y}_t = \mathbf{g}(\mathbf{x}_t, \mathbf{u}_t) + \mathbf{v}_t}_{\text{Observation Equation}} \quad \text{with} \quad g(\cdot) \sim GP(m_g(\cdot; \boldsymbol{\Phi}_g), k_g(\cdot, \cdot; \boldsymbol{\Theta}_g)).$$

ここで \mathbf{y}_t は離散時間 t における観測ベクトルであり、 \mathbf{x}_t は潜在的な状態ベクトルである。また、 \mathbf{e}_t は状態に関する共変数、 \mathbf{u}_t は観測の共変数であり、 \mathbf{v}_t と \mathbf{e}_t は状態と観測ノイズである。 $GP(m_g(\cdot; \boldsymbol{\Phi}_g), k_g(\cdot, \cdot; \boldsymbol{\Theta}_g))$ はガウス過程を表す。ガウス過程は平均関数 $m_g(\cdot; \boldsymbol{\Phi}_g)$ と共分散関数 $k_g(\cdot, \cdot; \boldsymbol{\Theta}_g)$ により特徴付けられる。これらの関数はその形が既知である、もしくはモデル選択により決定されるパラメトリックな関数である。本方法では、平均関数に“アンカーモデル”を仮定する。このアンカーモデルにより、問題に応じた線形／非線形な動的システムの構造を組み込むことができるようになる。また、状態と観測方程式のガウス過程の平均関数の事前分布として利用できる。

Monte Carlo Dynamic Classifier (MCDC) ツールには、次の二つのクラスのモンテカルロ法に基づく推

定アルゴリズムが実装されている。第一のクラスはパス空間モンテカルロ法 (Path-Space Monte Carlo; PSMC)、第二のクラスは粒子モンテカルロ法 (Particle Markov Chain Monte Carlo; PMCMC) に基づいている。MCDC ツールでは、状態と観測方程式に与えられる加法性ノイズが任意の分布に従うことを許容している。ガウス過程状態空間モデルについていくつかの先行研究があるが、一般にノイズは線形ガウスに制約されている。本ツールでは任意のノイズ分布を用いることができ、それらの方法の一般化と見なすことができる。このために、スパースグリッドテンソル積を用いてスプライン近似したガウス過程の状態／観測関数の値を、逐次モンテカルロフィルタリングにより計算する方法を開発・実装している。

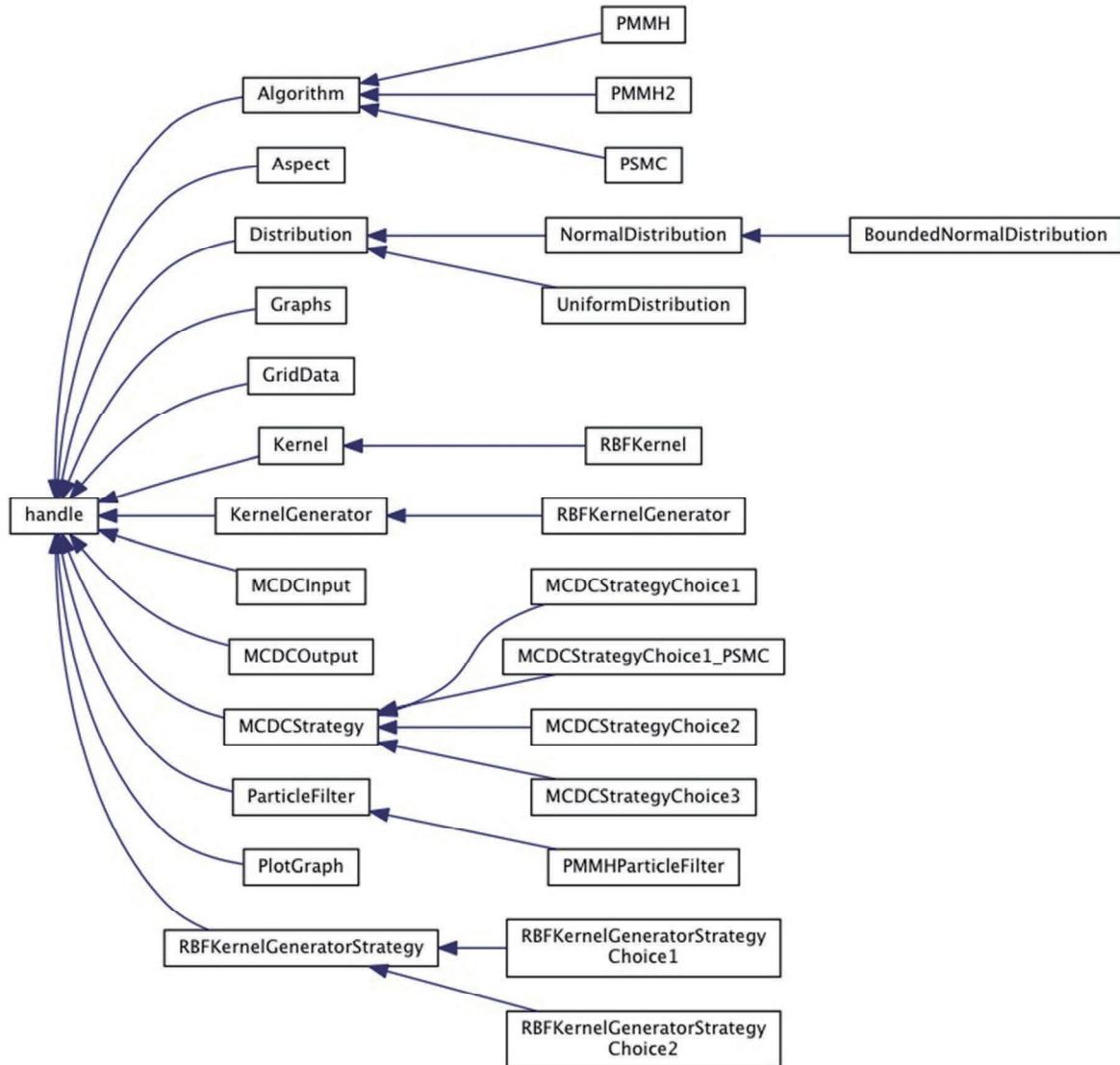


図 1 MCDC ツールのクラス階層図

2. MCDC ツールの概観

2.1. ソフトウェア構成

MCDC ツールはオブジェクト指向プログラミングに基づいて設計され、MATLAB 言語でコード化されている。図 1 に MCDC tool のクラス階層図を示す。本ツールの主要なクラスを次に示す。

MCDCInput and MCDCOutput :

MCDC ツールの入出力を扱うクラスである。

MCDCStrategy :

ガウス過程の平均関数・共分散関数、スプライン近似のためのスペースグリッド点、受理確率の計算手法を指定するためのクラスである。共分散関数を扱うために Kernel クラスと KernelGenerator クラスが用いられる。このクラスによりグラム行列の作成法(カーネル関数)が定義される。RBFKernelGeneratorStrategy クラスはカーネル関数のハイパーパラメータの推定法を決定するために利用される。MCDCStrategy クラスはガウス過程状態空間モデルの推定アルゴリズムを指定するための Algorithm クラスのインスタンスであるプライベート変数 algorithm を保持している。このクラスはまた、sequential importance resampling (SIR) を実行するための ParticleFilter クラスのインスタンスである プライベート変数 particle filter を保持している。

Graphs :

推定結果を可視化するためのクラスであり、2 次元グラフをプロットするための PlotGraph クラスである。

上記クラスで定義される関数の集合は以下の 3 つのコンポーネントに分類することができる。

学習コンポーネント :

モデル選択のための周辺尤度推定を事項するための関数群、すなわち、ガウス過程状態空間モデルのスペースグリッドテンソル積によるスプライン近似を用いた状態／観測関数の推定、状態の軌跡とフィルタ分布を推定するための関数群からなる。推定アルゴリズムとして、次の三つが用意されている。

PSMC :

未知の状態・観測方程式、状態系列、静的なモデルパラメータに由来する、全ての不確実性を統合した周辺尤度の推定量を得るためにパス空間モンテカルロ法を実行する。

PMCMC :

本アプローチでは、ガウス過程状態空間モデルに対応して、粒子マルコフ連鎖モンテカルロ法に基づいて事後分布からマルコフ連鎖サンプルを生成する。

PMCMC2 :

PMCMC の代替版であり、テンソル積スプラインを再構築して、より効率的な近似形で状態と観測関数を推定する。

テストコンポーネント :

観測データが与えられたときの分類問題を実行するための関数、およびガウス過程状態空間モデルのパラメータを推定するための関数群からなる。

グラフコンポーネント :

推定結果を可視化するための関数群からなる。可視化結果は Graph クラスが提供する関数群を用いて PDF ファイル形式で出力される。

3. Lorenz モデルによる状態系列の推定実験

Lorenz モデルによる状態系列の推定実験の結果について報告する。状態と観測関数は次の Lorenz モデルにより与える。

状態関数:

$$x_{(t,1)} = \sigma x_{(t-1,2)} \Delta t + x_{(t-1,1)} (1 - \sigma \Delta t) + \sqrt{q_1} \times v_{(t,1)} \quad (2)$$

$$x_{(t,2)} = x_{(t-1,1)} (\rho - y_{(t,1)}) \Delta t + x_{(t-1,2)} (1 - \Delta t) + \sqrt{q_2} \times v_{(t,2)} \quad (3)$$

観測関数 :

$$y_{(t,1)} = x_{(t-1,1)} x_{(t-1,2)} + y_{(t-1,1)} (1 - \beta) + \sqrt{r_1} \times n_{(t,1)} \quad (4)$$

$$y_{(t,2)} = \alpha x_{(t-1,1)} + \sqrt{r_2} \times n_{(t,2)} \quad (5)$$

$$y_{(t,3)} = \alpha x_{(t-1,2)} + \sqrt{r_3} \times n_{(t,3)} \quad (6)$$

表 1 ハイパーパラメータの設定

パラメータ名	説明	値
T	データ系列長	1000
σ	モデルパラメータ	10
ρ	モデルパラメータ	28
α	モデルパラメータ	1
β	モデルパラメータ	8/3
(q_1, q_2)	状態遷移関数に加えられるノイズの大きさ	(0.1 0.1)
(r_1, r_2, r_3)	観測関数に加えられるノイズの大きさ	(0.1 0.1 0.1)
Δt	時間間隔	0.01

表 2 Lorenz モデルの実験条件

説明	値
状態空間の次元数	2
状態遷移のカーネル関数	状態空間の第 1 次元を RBF カーネルに利用
観測のカーネル関数	状態空間の第 1 次元を RBF カーネルに利用
粒子数	500
MCMC の繰り返し回数	2000

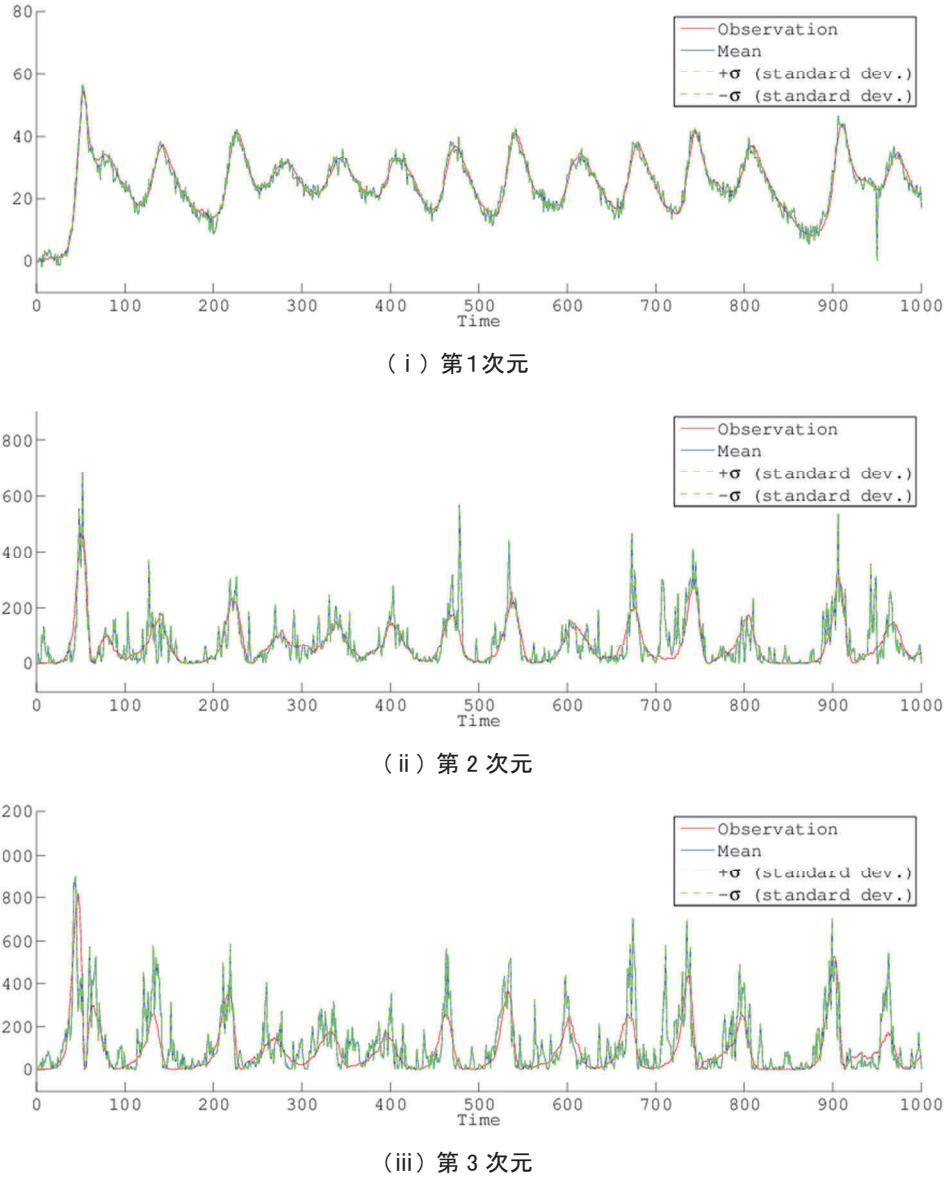


図 2 Lorenz モデルについて観測ベクトルの各次元ごとの真の周辺尤度と周辺尤度推定量との比較図

ただし σ, ρ, a, β は静的なモデルパラメータであり、各変数の定義は次の通りである；

$$\mathbf{x}_t = (x_{(t,1)}, x_{(t,2)}), \quad \mathbf{y}_t = (y_{(t,1)}, y_{(t,2)}, y_{(t,3)}), \quad \mathbf{v}_t = (v_{(t,1)}, v_{(t,2)}), \quad \mathbf{n}_t = (n_{(t,1)}, n_{(t,2)}, n_{(t,3)}).$$

変数 \mathbf{v}_t と \mathbf{n}_t が状態／観測方程式のノイズを表し、パラメータ q_1, q_2, r_1, r_2, r_3 はそれらノイズの振幅の大きさを制御する。

3.1. 実験条件

上記方程式のハイパーパラメータの設定を表 1 に、他の設定を表 2 に示す。

3.2. 実験結果

図 2 に各 t の観測で評価した局所周辺尤度の推定結果と PMCMC 法によるマルコフ連鎖周辺尤度の推定量を示す。Lorenz モデルの観測 $y_{(t,1)}$ は式 (3) に示す状態遷移の式に組み込まれる。つまり、状態関数は状態変数のみの関数として表現することはできない。MCDC ツールでは、このような場合に対応するために、追加の補助変数を導入し、状態変数に含めることができる。

4. モーションの分類実験

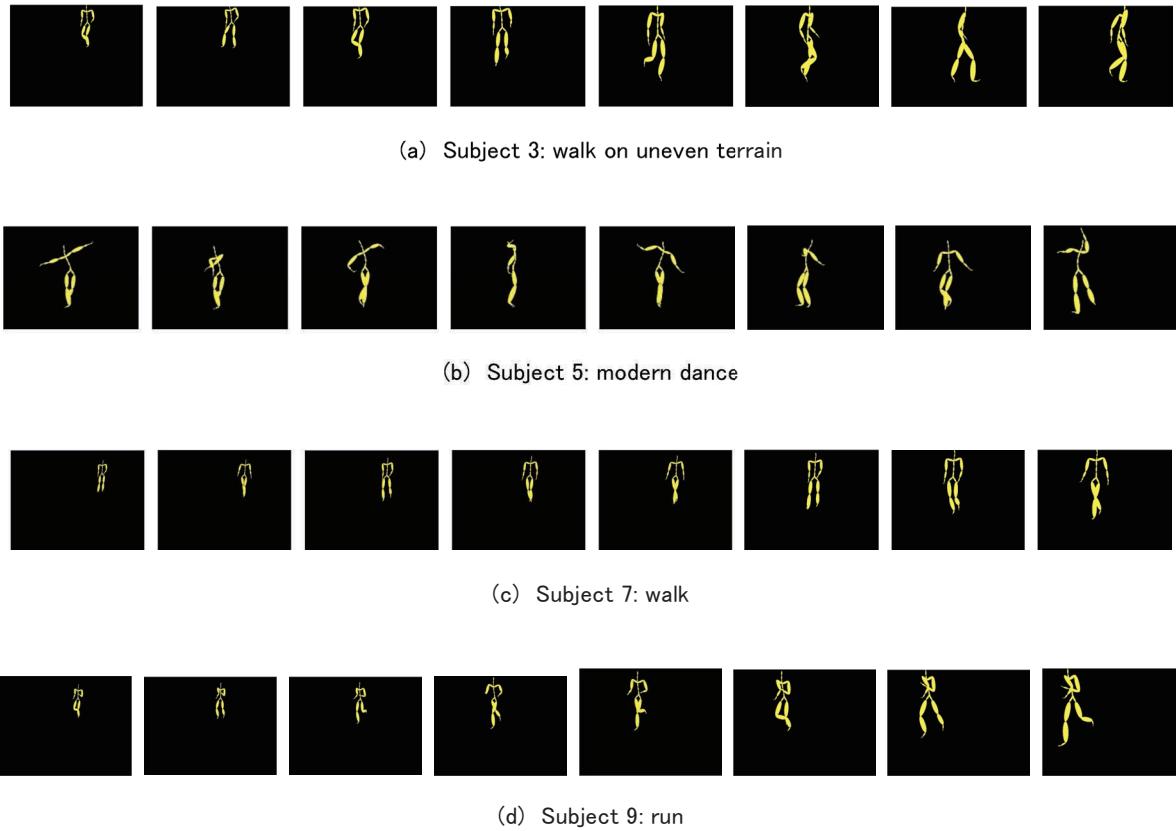


図 3 モーションキャプチャデータの例

(CMU Graphics Lab Motion Capture Database, <http://mocap.cs.cmu.edu> から)

表 3 モーションキャプチャデータの詳細

被験者番号	説明	試行数
3	walk on uneven terrain	4
5	modern dance	20
7	walk	12
9	run	12

CMU Graphics Lab Motion Capture Database を用いたモーションの分類実験の例を示す。このデータベースは 144 人分のデータを含む。実験では表 3 に示す 4 人の被験者のモーションデータを用いた。図 3 にその例を示す。被験者ごとに、複数回のモーションを試行している。各モーションについて、PMCMC2 法を用いてガウス過程状態空間モデルを推定した。評価時には leave-one-out 交差検定法を適用した。各モーションについて尤度を計算し、そのモーションを除く他の全てのモーションから計算した尤度と比較して、最も高い尤度を示すモーションのモデルを分類結果とした。

表 4 実験条件

説明	値の設定
状態空間の次元数	2
状態遷移の平均関数の初期値	各時刻で 0 の系列
状態遷移のカーネル関数	状態空間の第 1 次元を RBF カーネルに利用
観測のカーネル関数	状態空間の第 1 次元を RBF カーネルに利用
観測の平均関数の初期値	各時刻で 0 の系列
粒子数	500
MCMC の繰り返し回数	200

		Estimated subject number				Total	Accuracy (%)
		3	5	7	9		
Target subject number	3	1	2	1	0	4	25.0
	5	0	5	3	12	20	25.0
	7	0	1	7	4	12	58.3
	9	0	1	0	11	12	91.7

図 4 混同行列および分類精度

4.1. 実験条件

ここでは簡単に、一定速度を持ち、加速度 0 の直線運動を表すアンカーモデルを用いた。なお、使用データの大部分に対してこのアンカーモデルは不適切であり、一定速度の直線運動を反映するモーションについてのみ適している。このアンカーモデルを次に示す。

$$m_f(\mathbf{x}) = E[f(\cdot)] = A\mathbf{x} + B$$

$$m_g(\mathbf{y}) = E[g(\cdot)] = C\mathbf{x} + D$$

状態と観測関数のガウス過程事前分布として RBF カーネルを選択し、そのハイパーパラメータは学習時に PMCMC2 法により推定した。PMCMC2 法の設定を表 4 に示す。なお、実際には平均関数は未知な場合が多く、いろいろな形をとると考えられるが、MCDC ツールではアンカーモデルの関数形は任意に設定できる。

4.2. 分類結果

実験条件に従って学習したモデルデータを用いて、leave-one-out 交差検定法により性能を評価した。一定速度の構造という仮定に基づいて学習されたモデルは、異なる角加速度と非直線的な軌跡を伴う複雑なモーションには適さない。簡単な被験者 7 と 9 のデータには、ある程度の性能が得られることが期待される。

図 4 にクラス分類の混同行列と分類精度を合わせて示す。本実験ではまず、各モーションデータについ

て、PMCMC2 法を用いてモデルを推定した。その後、被験者の各モーションデータに対して、そのデータに対応するモデルを除外した全てのモデルから周辺尤度を計算し、その中から一番高い値を示すモデルの被験者番号を分類結果とした。図 4 より、特に被験者 7 と 9 について比較的高い性能が得られたことが分かる。また、被験者 7 と 9 のモーションは相対的に安定している一方で、被験者 3 と 5 のモーションは動的に大きく変化しており、線形ガウスの動的な構造に適さなかったと考える。なお、これらの結果は、一定速度ではないアンカーモデルを設定することで改善できると考える。

5. おわりに

MCDC ツールのドキュメント、ソースコード、実行例は web サイトから参照できる。

<http://www.ismvideo.org/NDE0UP/MCDCtool/en/index.html>

本ツールには今後も新たな機能を付加していく予定であり、それらの変更は上記 web サイトに反映する。