

平成 30 年 6 月 15 日現在

機関番号：12608

研究種目：基盤研究(B) (一般)

研究期間：2014～2017

課題番号：26280095

研究課題名(和文) ウェットGAの生物学実験データの分析に基づく高度並列型進化計算の実現

研究課題名(英文) Implementation of massive parallel Evolutionary Computation from Biological experiment data analysis obtained by wetGA

研究代表者

山村 雅幸 (Yamamura, Masayuki)

東京工業大学・情報理工学院・教授

研究者番号：00220442

交付決定額(研究期間全体)：(直接経費) 12,600,000円

研究成果の概要(和文)：生命にアイデアを得た進化計算の応用を通じて培われた探索戦略を、分子上に実装したウェットGAを提案し、タンパク質工学に応用してきた。本研究はこれらの成果を踏まえ、スパイラルの次の段階として計算モデルを洗練化し、実装を試みた。予想したほど並列計算の効率が上がらないという困難が観察された。集団サイズと、最適解を発見するまでの世代数とのトレードオフについて、巡回セールスマン問題と関数最適化問題を用いて、より詳細に調べた。解くべき最適化問題が簡単すぎると、並列度は無駄になることがわかった。

研究成果の概要(英文)：Evolutionary Computation is an optimization search technique inspired by Biological Evolution. We inversely developed a molecular implementation of Evolutionary Computation, called wet GA, and applied it in Protein Engineering. This research shows the next stage of this Biology / Computer spiral. We improved conventional Evolutionary Computation models with the knowledge obtained by wet GA. We found the fact that can be hard to achieve expected performance even with massive parallelism. There is a trade-off between the population size, i.e. the degree of parallelism, and the generation number, i.e. the time to find the optimum. We examined the trade-off by using travelling salesman problems and typical function optimization problems. The results show that too much parallelism will become wasteful when the problem has a simple search landscape.

研究分野：DNAコンピューティング

キーワード：ウェットGA 超並列進化計算 WetTDGA 分子進化 NK地形 疑似適応度 パラメータ最適化 細胞シミュレーション

1. 研究開始当初の背景

生命の「進化」という現象を工学的に利用する試みには、情報科学上の進化計算（遺伝的アルゴリズム：GA）、生命科学上のタンパク質工学（分子進化学）がある。これらは「変異と選択による最適解の発見」という大雑把な骨格は共有しているが、利用可能な資源・手段の制約や歴史的経緯などによって異なる探索戦略を発達させてきた。

進化計算は初期のナイーブな発想のレベルを抜け、確率論に基礎を置いた信頼性の高い近似最適手法となった。探索は(1)ランダムな初期集団から出発する。変異は(2)複数親の交叉を主に用いる。(3)適応度評価は自由にプログラムできる。一方、直列計算機が想定されており、比較的少ない集団サイズで多くの世代交代を重ねる(4)狭くて深い探索戦略の実装が続けられてきた。

タンパク質工学は生物学実験室ではごく普通の実験技術である。進化計算と対照的に、探索はすでに機能している単独のタンパク質、すなわち(1)単一の局所解から出発する。変異は DNA の実験操作時に自然に導入される。交叉でシャッフルをかけることもあるが、(2)単独親の突然変異が基本である。実験技術として(3)適応度評価は困難であり、定量化できることはめったにない。一方、数多くの変異体を同時並列的に培養することは容易だが、世代交代を重ねることはほとんどなく、(4)広くて浅い探索戦略の実装が普及している。

進化計算における探索の進捗状況は、現世代の個体分布として保持されている。単独親の突然変異よりも、複数親の交叉の方が、個体分布に忠実なサンプリングができる。連続関数最適化では、多峰性の困難な地形において他の近似最適手法より優れた性能を見せるまでになっている。この経験から見ると、タンパク質工学の探索戦略は選択圧が高すぎて局所最適解に初期収束してしまうように見受けられる。実際、タンパク質工学では富士山型の単峰性の地形が想定されており、

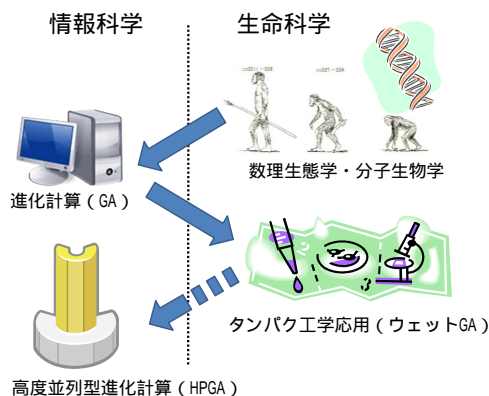


図1 ウェットGAによる異分野融合スパイラル

地球生命の生体分子はこのような地形に沿って進化してきたとまで唱えられている。理論解析でも単独親の突然変異のみの進化戦略 (ES) のモデルが利用される。

近年の情報科学と生命科学の両分野の技術革新によって、従来の探索戦略を方向づけてきた制約は絶対的条件ではなくなりつつある。これらの暗黙の前提条件の見直しによって、より効率の高い進化計算、より困難な分子進化といった異分野融合スパイラルの次の段階に入る条件が整ったと考える。

2. 研究の目的

これまでに tyrRS の基質改変の適応度地形の特徴をモデル化し、探索が1つの局所解からはじめられる進化計算の効果的な世代交代モデルを提案して、理論設計に従って実際に tyrRS を進化させた[1][2]。各世代の tyrRS 変異体の遺伝子シーケンスの全解読データが得られているので、そこからエントロピーの推移を計算し、モデルを洗練化している。本研究では、特定の世代を選んで tyrRS 変異体を再度進化させ、パラメータ感度の意味でのモデルの正しさを検証する。さらに、これらの知見を発展させて、高度並列アーキテクチャーを前提とした新しい進化計算を実装する。

3. 研究の方法

研究計画は計算的側面（理論および実装）と実験的側面（遺伝子組換え実験）の2つからなる。坂本らによる tyrRS の基質改変実験の結果[1][2]について、すでに全世代の適応度分布およびシーケンス結果が得られている。これらの分析をもとにして、計算面では、染谷らによる WetTDGA の枠組み[1][3]に基づいてモデルを構築し、実験データとの適合を調べる。実験面では、坂本らの実験のうち、初期・中期・終期から特徴的な世代を選び、そこから再度進化させてパラメータ感度解析に役立てる。最終的に高度並列化進化計算の理論としてまとめあげた上で、導入予定の計算サーバー上でデモンストレーション実装する。

【平成26年度】

(1)WetTDGA の最適化能力の理論解析
 (1a)最適化としての困難さを表現する新しいモデルの提案と解析：一般に、野生型遺伝子に加えられた変異がどのようにタンパクの機能に影響を与えるか、という現象のモデル化は困難である。本研究では、個々の現象の精密なモデル化のかわりに最適化の困難さに関わる地形の複雑さをモデル化する。タンパク質工学でよく用いられる NK 地形モデルは、tyrRS の基質改変のような大域的な多峰性の困難な最適化問題の地形の表現には不相当であると考え、新しいモデルを提案する。
 (1b)tyrRS の基質改変実験向け最適パラメータの設計：提案された世代交代モデルから、適応度分布に応じたサンプル個数など、実験

向けの最適パラメータを決定する。また、比較用としてネガティブなパラメータ設定も提案し、性能悪化の程度を見積もる。

(2) tyrRS の基質改変のテストベッドを用いた確認実験

(2a) 各世代のシーケンス全解読による集団分布の確認実験：坂本らによってすでに 12 世代分の個体群を得ている。理論解析とそれに基づく世代交代手順の適切さを証明するために、世代交代によって集団中の変異の分布がどのように変化したのかを、全個体の tyrRS 変異体の遺伝子シーケンスの解読結果に基づき分析する。遺伝子レベルでのエントロピーの変化が、理論解析と合致するかどうかを確認する。

(2b) 比較用世代交代バリエーションによる進化実験：成功例が偶然の産物ではないことを証明するために、理論解析から性能の悪化や失敗が予想される手順に基づいて再進化実験を行い、同様に世代交代によって集団中の変異の分布がどのように変化したのかを、全個体のシーケンスによって評価する。

【平成 27 年度以降】

(1) WetTDGA の最適化能力の理論解析：提案した地形モデルを tyrRS の基質改変以外にも使えるように一般化する。一般化されたモデルから実験手順やパラメータの見積り等の方法論を確立する。

(2) 高度並列型進化計算の実現：ここまでで得られた結果を敷衍して、マルチコアやネットワーク上に分散されたクラウドなどの高度な並列性が安価に実現できる計算機アーキテクチャーを前提として、実装コスト・計算量・解の良否の総合的な観点から効果的に実現できる、高度並列型進化計算の枠組みとその性能の理論解析にまとめ上げ、計算機上に実装する。

4. 研究成果

(1) tyrRS 変異体シーケンスの解析：当初計画のうち、全世代の tyrRS 変異体の遺伝子シーケンスの全解読は終了し、各世代のエントロピー推移とモデルとの適合を検討した。追加のウェット実験の設計を行った。特にシーケンスの全解読データの解析には注目すべき知見を得て、新しい進化計算の理論的基盤を考案した。この枠組みについては、さらに詳細な検証が必要である。

最終目標である高度並列 GA の実装のために、ウェット実験で得られたデータからデータマイニング手法に基づいて、ウェット実験の適応度関数の近似関数を求めた。その結果、複数の近似関数候補が見つかり、それによって、高度並列 GA の実装形態が複数考えられることがわかった。

(2) 進化計算の並列度と世代数の有効性トレードオフの発見：計算機上の進化計算で培われた探索戦略をタンパク工学に応用したウェット GA の実用から得られた遺伝子シーケンスのデータ群を解析し、理論と

の合致を調べた。解析の途上で、当初予想した並列度と世代数の有効性トレードオフが、解くべき最適化問題の難易度に依存することを発見した。

(3) 問題難易度に依存する並列度 / 世代数トレードオフの解析：理論解析の途上で、当初予想したほど並列計算の効率が上がらないという困難がシミュレーションによって観察された。進化計算の並列度すなわち集団サイズと、最適解を発見するまでの世代数とのトレードオフについて、代表的なベンチマーク問題である巡回セールスマン問題と関数最適化問題を用いて、より詳細に調べた。解くべき最適化問題が簡単すぎると、並列度は無駄になることがわかった。

(4) 実問題アプリケーション：関連課題として、開発した進化計算を免疫細胞の動的応答のモデリングにおけるパラメータ最適化に応用した。それまで知られていなかった新しいオフセット要素の存在を示唆する結果を得た。

< 引用文献 >

[1] Sakamoto K., Yamamura M., Someya H.: Toward "Wet" Implementation of Genetic Algorithm for Protein Engineering, Proceedings of 10th International meeting on DNA based computers (DNA10), 416-425 (2004).

[2] Irahara, F., Oki, K., Kobayashi, T., Ohno, S., Yokogawa, T., Nishikawa, K., Yokoyama, S., Sakamoto, K.: Functional replacement of the endogenous tyrosyl-tRNA synthetase tRNA^{Tyr} pair by the archaeal tyrosine pair in Escherichia coli for genetic code expansion, Nucleic Acids Res. 2010 Jun;38(11):3682-91. doi: 10.1093/nar/gkq080. Epub 2010 Feb 16.

[3] 染谷博司, 山村雅幸, 坂本健作: "分子計算のための一点から開始される探索法", 人工知能学会論文誌, Vol.22, No.4, pp.405-415, 5 月 2007. #DOI: 10.1527/tjsai.22.405.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 3 件)

Ishimatsu K, Hata T, Mochizuki A, Sekine R, Yamamura M, Kiga D, General applicability of synthetic gene-overexpression for cell-type ratio control via reprogramming, ACS synthetic biology, 有, 3(9), 2014, 638-644, 10.1021/sb400102w

Shoya Yasuda, Yang Zhou, Yanqing Wang, Masayuki Yamamura, Ji-Yang Wang, Vol. 7, A model integrating tonic and activated BCR signals to

predict the survival of primary B cells, 2017 年, Scientific Reports, 14888, 10.1038/s41598-017-13993-x, 無

Yusuke Yanagita, Masayuki Yamamura, Gradient Masking is a Type of Overfitting, 2017 年, Gradient Masking is a Type of Overfitting, 無

〔学会発表〕(計 7 件)

K. Komiya, C. Noda, M. Komori, K. Dong, T. Takenaka, T. Enomoto, T. Yoshimura, M. Yamamura, Construction of an isothermal DNA amplification reaction system for simplified nucleic acid test, 第 10 回 バイオ関連化学シンポジウム, 2016 年 9 月 7 日 ~ 2016 年 9 月 9 日, 石川県立音楽堂

小宮健, 若林優太, 山村雅幸, 翻訳機構の進化過程解明に向けたリボソームの立体構造比較, 「細胞を創る」研究会 9.0, 2016 年 11 月 21 日 ~ 2016 年 11 月 22 日, 早稲田大学

若林優太, 小宮健, 山村雅幸, 好熱性細菌を対象としたリボソームの立体構造比較, 極限環境生物学会 2016 年度 (第 17 回) 年会, 2016 年 11 月 25 日 ~ 2016 年 11 月 26 日, 東京工業大学

Shoya Yasuda, Yang Zhou, Ji-Yang Wang, Masayuki Yamamura, B 細胞の生存を制御する内因性および外因性シグナルの数理解析の試み, 第 89 回日本生化学会大会, 2016 年 09 月 25 日 ~ 2016 年 09 月 27 日, 仙台国際センター
Shoya Yasuda, Yang Zhou, Yanqing Wang, Masayuki Yamamura, JiYang Wang, Mathematical modeling of intrinsic and extrinsic signals that regulate B cell survival, 第 45 回日本免疫学会学術集会, 2016 年 12 月 05 日 ~ 2016 年 12 月 07 日, 沖縄コンベンションセンター

Shoya Yasuda, Yang Zhou, Ji-Yang Wang, Masayuki Yamamura, B 細胞の生存を制御する BAFF 受容体シグナルの数理解析の試み, 生命科学系学会合同年次大会(ConBio2017), 2017 年

Shoya Yasuda, Yang Zhou, Masayuki Yamamura, Ji-Yang Wang, Mathematical modeling of BAFFR-mediated B cell survival signal, 第 46 回日本免疫学会学術集会, 2017 年

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 0 件)

取得状況 (計 0 件)

〔その他〕

ホームページ等: 該当なし

6. 研究組織

(1) 研究代表者

山村 雅幸 (YAMAMURA, Masayuki)
東京工業大学・情報理工学院・教授
研究者番号: 00220442

(3) 連携研究者

坂本 健作 (SAKAMOTO, Kensaku)
独立行政法人理化学研究所・生命機能科学
研究センター・非天然アミノ酸技術研究チ
ーム チームリーダー
研究者番号: 50240685

染谷 博司 (SOMEYA, Hiroshi)
東海大学・情報理工学部・准教授
研究者番号: 00333518

小宮 健 (Komiya, Ken)
東京工業大学・情報理工学院・助教
研究者番号: 20396790