

令和元年6月15日現在

機関番号：34416

研究種目：基盤研究(C)（一般）

研究期間：2015～2018

課題番号：15K00402

研究課題名（和文）インタラクティブな大規模ベイジアンネットワーク推定法の提案と生物データへの応用

研究課題名（英文）Proposal of interactive large scale Bayesian network estimation method and its application to biological data

研究代表者

竹中 要一（takenaka, yoichi）

関西大学・総合情報学部・教授

研究者番号：00324830

交付決定額（研究期間全体）：（直接経費） 3,600,000円

研究成果の概要（和文）：データ解析をするにあたり事象間の原因と結果を表す因果関係は、単に事象間の変化が類似している事象を表す相関関係よりも重要である。この因果関係を表現するモデル及び因果関係の推定手法として、ベイジアンネットワークは小規模なデータ解析に大きく貢献してきた。ベイジアンネットワークは因果関係を条件付き確率で表した確率モデルであり、事象に対応する頂点間の有向辺で因果関係を表す非循環有向グラフである。

本研究では、ベイジアンネットワークが推定可能な事象数を拡大する事、同一事象を対象とし異なる情報源から観測されたデータを比較する方法を提案した。そして有効性を生物学分野の実データを用いた検証を通じて明らかとした。

研究成果の学術的意義や社会的意義

ベイジアンネットワークによる因果関係解析は、生物学、医学、薬学に止まらず医師の診断、イメージ認識、言語認識でも用いられる汎用的な手法である。各分野において今後もデータ産生量が増大し続ける事は確実であろう。そのため、因果関係を推定可能な事象数を拡大可能な手法、および観測データ比較を可能とする手法の需要は今後も拡大し続けるであろう。そのため、本研究の成果は生物学に止まらず、他分野データの解析においても新しい知見を獲得するための手段として活躍する事が期待される。

研究成果の概要（英文）：In data analysis, causality that indicates cause and effect between events is more important than correlation that simply indicates that changes between events are similar. Bayesian networks have greatly contributed to small-scale data analysis as a model expressing the causality and a method of estimating the causality. The Bayesian network is a probabilistic model representing causality with conditional probability, and is a non-cyclic directed graph representing causality with directed edges between vertices corresponding to events.

In this study, we proposed a method to expand the number of events that Bayesian networks can estimate and to compare data observed from different sources for the same event. And the effectiveness was clarified through verification using real data in the field of biology.

研究分野：生物情報学

キーワード：生物情報学 ベイジアンネットワーク 遺伝子制御ネットワーク 一塩基多型 大規模データ 因果関係

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

データ解析をするにあたり事象間の原因と結果を表す因果関係は、単に事象間の変化が類似している事象を表す相関関係よりも重要である。この因果関係を表現するモデル及び因果関係の推定手法として、ベイジアンネットワークは小規模なデータ解析に大きく貢献してきた。ベイジアンネットワークは因果関係を条件付き確率で表した確率モデルであり、事象に対応する頂点間の有向辺で因果関係を表す非循環有向グラフである。ベイジアンネットワークによる因果関係解析の歴史は古く、1980年代より医師の診断、イメージ認識、言語認識へ用いられてきた。その後も利用される分野は広がりつづけ、データ産生量が増大している生物学、医学、薬学の各分野において効果的な手法として利用され続けている。ベイジアンネットワークの問題点として、因果関係の推定に必要な計算量が事象の数に対して指数関数的に増加する事、すなわち実用的な時間で因果関係の推定を可能とする事象の数が少ない事が挙げられる。そのために分枝限定法や各種ヒューリスティクス技法が提案されてきたが、それでもなお数千事象間の因果関係の推定にはスパコンを必要としていた。そこではヒューリスティクス技法とブートストラップを主体とする、スパコンの計算速度に依存した、美しくないアルゴリズムを用いていた。また、生物学データのような事象数が多いデータを対象とする因果関係解析としては、数千事象はまだまだ少なすぎた。ヒトの遺伝子数が数万、各遺伝子のアイソフォームが10個弱、SNPが10~100個。そして父方と母方から1対ずつ受取る事による組み合わせ(ハプロタイプ)も併せると、事象数は容易に100万以上に膨れあがる事になる。さらにゲノムのメチル化等のデータもここに加わる事になる。このためには、スパコンの計算速度に依存したアルゴリズムを根本的に変更する必要がある、かつ利用者の利便性を考慮して研究室レベルで準備可能な計算サーバー、または個人利用のパソコンでの実行を可能とする必要があった。

ベイジアンネットワーク推定には他の問題点も存在していた。ベイジアンネットワークで表現される因果関係を計算機で推定する場合、基本的にある基準で最も優れたものを出力する。基準としては、ベイズ情報量基準や最小記述長がしばしば用いられるが、この基準で最も優れたものは必ずしも現実と整合しない。整合しない理由は、入力されるデータが少なかったり偏りがあったり、偶然の産物だったり様々ではあるが、少なくとも生物情報学分野では推定された因果関係を過度に信頼する事はできない。そのため出力された因果関係を基に他のデータと照らし合わせたりしながら試行錯誤的に修正を行っていく事になる。この試行錯誤に対応可能なアルゴリズム、および試行錯誤を可能とする高速化が必要であった。

2. 研究の目的

上述の背景に記載した問題点の解消、あるいは緩和を目的としたアルゴリズム等の提案を行う。提案手法の有効性は各種生物学データを用いて検証する事を研究の目的とした。

ベイジアンネットワークはクラス NP (non-deterministic Polynomial) に属するため、高速計算を実現する多くのヒューリスティクスが提案されてきた。しかし実際に最も多く用いられる手法はランダムな初期ネットワークからの局所改善を繰り返す方法である。この方法はこの方法は並列化には適しているものの、事象数の増加に従い探索される空間の割合が逆指数関数的に減少する。従って事象数の増大には適していない。そこで本研究では、局所的な改善を繰り返す方法ではなく、小規模だが網羅的に組み合わせた事象集合ごとに推定したベイジアンネットワークに基づき、一度に全事象の因果関係を推定する手法を用いる事で、高精度かつ高速に事象数の大きなベイジアンネットワークを推定する方法を用いる。またこの手法の小規模ベイジアンネットワークの組み合わせ部分の計算量を小さくする事により、試行錯誤を可能とするアルゴリズムとする。また、この試行錯誤によって疾患などの目的変数を説明する事象集合

の推定も可能となる。これらは、遺伝子の時系列発現プロファイル、ゲノムワイドの一塩基多型およびメチル化頻度データを用いる事でその有効性を検証する。

3. 研究の方法

ベイジアンネットワークによる因果関係推定を可能とする事象数(N)は10万程度を設定した。これを実現するためには計算量がNの多項式、それも最高次数を小さい値に抑える必要がある。そこで本研究で用いるアルゴリズムの計算量が $O(N^3)$ で収まるようにする。計算量が $O(N^3)$ となる律速部分は因果関係推定部になる。ここでは $O(N^3)$ 個の独立した部分問題を計算している。そのため多数のCPUコア、GPUを使った並列実装によってさらなる高速化を図る事ができる。

4. 研究成果

提案したベイジアンネットワーク推定手法に対し、各種生物学データを適用した結果、その有効性を明らかとしている。雑誌論文[1][6]および学会発表[19]では、マウスの間葉系幹細胞から脂肪細胞および骨芽細胞への分化過程で採取した遺伝子発現プロファイルを使い、ベイジアンネットワークで遺伝子制御ネットワークを推定した上で、その比較解析を行っている。この解析により2種類のデータを試行錯誤的に比較解析する事の有効性を明らかにしている。また、上記論文では、時系列プロファイルに含まれる環境条件を複数集合に分離する方法の有効性も明らかにしている。大腸菌が培地に含まれるブドウ糖を消費し尽くした後、栄養源を乳糖に変更する過程を観測した遺伝子発現プロファイルがある。ベイジアンネットワークを用いる事により生存環境が大幅に変化する時点、すなわちブドウ糖が枯渇する時点を同定する事ができる事を明らかにした。また、学会発表[18][42]では、因果関係推定を行う事象数を効率的に削減する事を可能とする手法を提案し、乳がんに関連するヒトSNPを高精度に選択する事が可能である事を示している。

5. 主な発表論文等

〔雑誌論文〕(計6件)

- [1] [Yoichi Takenaka](#), Kazuma Mikami, Shigeto Seno, Hideo Matsuda, Automated transition analysis of activated gene regulation during diauxic nutrient shift in Escherichia coli and adipocyte differentiation in mouse cells, BMC bioinformatics, vol 19(suppl 4):89 (2018/5)
DOI: 10.1186/s12859-018-2072-y 査読あり
- [2] Viacheslav Kapilevich, Shigeto Seno, Hideo Matsuda, [Yoichi Takenaka](#) Chromatin 3D reconstruction from chromosomal contacts using a genetic algorithm IEEE/ACM Transactions on Computational Biology and Bioinformatics, DOI: 10.1109/TCBB.2018.2814995 (2018/3) 査読あり
- [3] Tazro Ohta, Takeshi Kawashima, Natsuko O. Shinozaki, Akito Dobashi, Satoshi Hiraoka, Tatsuhiko Hoshino, Keiichi Kanno, Takafumi Kataoka, Shuichi Kawashima, Motomu Matsui, Wataru Nemoto, Suguru Nishijima, Natsuki Suganuma, Haruo Suzuki, Y-h. Taguchi, [Yoichi Takenaka](#), Yosuke Tanigawa, Momoka Tsuneyoshi, Kazutoshi Yoshitake, Yukuto Sato, Riu Yamashita, Kazuharu Arakawa, and Wataru Iwasaki. (23 員数中 16 番目) Collaborative environmental DNA sampling from petal surfaces of flowering cherry Cerasus x yedoensis 'Somei-yoshino' across the Japanese archipelago. Journal of Plant Research, Jul;131(4), p.709-717, doi: 10.1007/s10265-018-1017-x. (2018/2) 査読あり
- [4] Babak Ehteshami Bejnordi, et. al. ([Y. Takenaka](#) は 61 員数中 50 番目), Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer, The Journal of the American Medical Association, Vol.318, No.22, pp.2199-2210, doi: 10.1001/jama.2017.14585 (December, 2017). 査読あり
- [5] Mikio WATANABE, [Yoichi Takenaka](#), Chika Honda, Yoshinori Iwatani, Genotype-based epigenetic differences in monozygotic twins discordant for positive anti-thyroglobulin autoantibodies, Thyroid, 28(1): 110-123. (2017/01)
DOI: 10.1089/thy.2017.0273 査読あり

- [6] Yoichi Takenaka, Shigeto Seno, Hideo Matsuda
Detecting shifts in gene regulatory networks during time-course experiments at single-time-point temporal resolution
J. Bioinformatics and Computational Biology, vol. 13 Issue 5,
doi: 10.1142/S0219720015430027, 全 18 頁 (2015-9) 査読あり
〔学会発表〕(計 42 件)
- [1] 仲川勇二, 花田良子, 竹中要一, 岩田員典, 回帰における多重共線性の全整数最適化アプローチ日本オペレーションズ・リサーチ学会 2019 年 春季研究発表会 2-B-10, (千葉工業大学) (2019/3/15)
- [2] Yoichi Takenaka, Transition of regulatory force toward the gene expressions during osteoblast cell differentiation, Pacific Symposium on Biocomputing (PSB) 2019, HI, USA, (2019-1-5)
- [3] 渡邊幹夫, 竹中要一, 本多智佳, 大阪ツインリサーチグループ, 岩谷良則, TgAb 出現が一致しない一卵性双生児ペアを対象とした, 自己抗体出現に関するエピゲノム要因の解析, 第 61 回日本甲状腺学会学術集会 (川越) (2018/11/23) (名古屋大学)
- [4] 渡邊幹夫, 竹中要一, 本多智佳, 大阪ツインリサーチグループ, 岩谷良則, 自己抗体の出現が一致しない一卵性双生児を対象とした, 自己抗体出現に影響する遺伝的要因とエピゲノム要因の解析, 第 46 回日本臨床免疫学会総会 (軽井沢) (2018/11/8)
- [5] Nakagawa Y., Hanada Y., Takenaka, Y., Edirishinghe C., Nonconvex Discrete Optimization Approach for High Dimensional Regression, INFORMS annual meeting 2018, Phoenix, AZ, USA (2018/11/07)
- [6] Watanabe, M., Takenaka, Y., Honda, C., Osaka Twin Research Group, Iwatani, Y., Genotype-based epigenetic factors in identical twins discordant for positive TgAb. AACCC Annual Meeting & Clinical Lab Expo 2018, Chicago, USA (2018/7/31)
- [7] 渡邊幹夫, 竹中要一, 本多智佳, 大阪ツインリサーチグループ, 岩谷良則, 自己抗体の出現が不一致な一卵性双生児における, 個体の遺伝背景を考慮した DNA メチル化率の解析, 第 12 回日本エピジェネティクス研究会年会 (札幌) (2018/5/24)
- [8] 尾崎律子, 渡邊幹夫, 竹中要一, 本多智佳, 富澤理恵, 岩谷良則, 一卵性双生児を対象とした HbA1c 値に影響するゲノムおよびエピゲノム因子の探索, 日本双生児研究学会 第 32 回学術講演会 演題 8, 大阪大学 (2018/1/27)
- [9] Yoichi Takenaka, Kazuma Mikami, Shigeto Seno, Hideo Matsuda, Automated transition analysis of activated gene regulation during diauxic nutrient shift in Escherichia coli and adipocyte differentiation in mouse cells, The 16th Asia Pacific Bioinformatics Conference (APBC2018), Yokohama Japan, (2018-01-15)
- [10] Kazumasa Saito, Shigeto Seno, Hideo Matsuda and Yoichi Takenaka, Missing value imputation for gene expression data with Deep Learning The 16th Asia Pacific Bioinformatics Conference, A9, Yokohama, Japan (2018/1/15-17)
- [11] Meya Cho, Shigeto Seno, Hideo Matsuda and Yoichi Takenaka, Functional analysis of metagenome using composition-based method, The 16th Asia Pacific Bioinformatics Conference, A18, Yokohama, Japan (2018/1/15-17)
- [12] Viacheslav Kapilevich, Shigeto Seno, Hideo Matsuda, Yoichi Takenaka, Chromatin 3D reconstruction from chromosomal contacts using a genetic algorithm, The 28th International Conference on Genome Informatics (GIW/BIOINFO 2017), Seoul Korea, Oct. 31- Nov.3 (2017-11)
- [13] Watanabe, M., Takenaka, Y., Honda, C., Osaka Twin Research Group, Iwatani, Y., Genotype specific for monozygotic twins discordant for anti-tyroglobulin autoantibody (TgAb) positivity. The 16th Congress of the International Society for Twin Studies (ISTS), Madrid, Spain (2017/11/16-18)
- [14] Ozaki, R., Watanabe, M., Takenaka, Y., Honda, C., Tomizaw, R., Iwatani, Y., Genetic and epigenetic factors influencing HbA1c levels in monozygotic twins. The 29th World Association of Societies of Pathology and Laboratory Medicine (WASPALM), Kyoto, (2017/11/17)
- [15] 嶋田彩人, 繁田浩功, 瀬尾茂人, 池田わたる, 竹中要一, 松田秀雄, 蛍光顕微鏡画像のためのグラフカットによる毛細血管の抽出手法, 画像の認識・理解シンポジウム (MIRU2017), PS1-65, (2017/8)
- [16] 尾崎律子, 渡邊幹夫, 竹中要一, 本多智佳, 富澤理恵, 大阪ツインリサーチグループ, 岩谷良則, 一卵性双生児を対象とした HbA1c 値に影響する遺伝因子と環境因子の探索, 57 回日本臨床化学会年次学術集会 (札幌) (2017/10/7)
- [17] Shigeto Seno, Yoichi Takenaka, Hideo Matsuda, Clustering of cells for single cell analysis using Latent Dirichlet Allocation, Mitsuhiro Eto, The 25th conference on Intelligent Systems for Molecular Biology and the 16th European Conference on Computation Biology (ISMB/ECCB 2017), B-071, Prague, Czech Republic (2017/7)
- [18] Mikio WATANABE, Yoichi Takenaka, Chika Honda, Yoshinori Iwatani, Genetic background

- specific for TgAb discordant twins, 47th Behavioral Genetics Association Annual Meeting, Oslo (2017/6/29)
- [19] 大田達郎, 川島武士, 土橋映仁, 平岡聡史, 星野辰彦, 菅野圭一, 片岡剛文, 川島秀一, 松井求, 根本航, 西嶋傑, 菅沼名津季, 鈴木治夫, 田口善弘, 竹中要一, 谷川洋介, 恒吉桃香, 吉武和敏, 荒川和晴, 山下理宇, 岩崎渉, お花見メタゲノムプロジェクト結果報告 NGS 現場の会 第五回研究会 2017-5-23/24, 仙台国際センター (2017-5)
- [20] 江藤充宏, 瀬尾茂人, 竹中要一, 松田秀雄, 一細胞解析のための潜在的ディリクレ配分法を用いた遺伝子群の機能予測手法, NGS 現場の会 第五回研究会 2017-5-23/24 仙台国際センター (2017-5)
- [21] 伊藤 澄美, 瀬尾 茂人, 繁田 浩功, 菊田 順一, 竹中 要一, 石井 優, 松田 秀雄, 細胞個体の移動傾向の違いを考慮したグローバルデータアソシエーションを用いた細胞追跡手法, 第 112 回数理解モデル化と問題解決研究発表会 2017-MPS-112-6 (2017-2)
- [22] 藪崎 隼人, 瀬尾 茂人, 竹中 要一, 後藤 剛, 河田 照雄, 松田 秀雄, 畳み込みニューラルネットワークを用いた脂肪細胞セグメンテーションにおける分割精度改善手法の提案 第 112 回数理解モデル化と問題解決研究発表会 2017-MPS-112-27 (2017-2)
- [23] Yoshiaki Sota, Shigeto Seno, Yoichi Takenaka, Shinzaburo Noguchi, Hideo Matsuda, Comparative analysis of transformation methods for gene expression profiles in breast cancer datasets, IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE 2016), Splendor Hotel, Taichung, Taiwan, pp.328-333, doi: 10.1109/BIBE.2016.51 (2016-11-1)
- [24] Yoshiaki Sota, Shigeto Seno, Yoichi Takenaka, Shinzaburo Noguchi, Hideo Matsuda, Comparative analysis of transformation methods for gene expression profiles in breast cancer datasets, IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE2016), pp.328-333, Taichung, Taiwan (2016/10)
- [25] Hideo Matsuda, Junko Tsuda, Shigeto Seno, Yoichi Takenaka, A method for reducing false positives of target genes of microRNAs by combining two prediction tools, International Conference on Intelligent Systems for Molecular Biology (ISMB2016), No.049, Swan and Dolphin Hotel, Orlando, Florida, USA (2016-7).
- [26] Fuki Iwasaki, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda, An accurate estimation method for expression profiles of novel transcripts from RNA-Seq data, International Conference on Intelligent Systems for Molecular Biology (ISMB2016), No.N24, Swan and Dolphin Hotel, Orlando, Florida, USA (2016-7).
- [27] 松原周平, 瀬尾茂人, 西澤志乃, 菊田順一, 竹中要一, 石井優, 松田秀雄, オプティカルフローとウェーブレット解析を用いた白血球の動態の解析手法, バイオイメージ・インフォマティクスワークショップ 2016, P03, 大阪大学銀杏会館, 大阪 (2016-6)
- [28] 渡邊誓旅, 瀬尾茂人, 竹中要一, 松田秀雄, 画像処理とディープラーニングの手法を用いた病理診断支援, バイオイメージ・インフォマティクスワークショップ 2016, P23, 大阪大学銀杏会館, 大阪 (2016-6)
- [29] Kazumasa Saito, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda Bayesian Network Inference from gene expression profiles with a small number of samples, The 20th Annual International Conference on Research in Computational Molecular Biology, RECOMB2016, Poster-50 (2016-4)
- [30] Yoichi Takenaka, Shigeto Seno, Hideo Matsuda, Dynamic Analysis of Gene Regulations using Leaving-One-Out Expression Profile, The 20th Annual International Conference on Research in Computational Molecular Biology, RECOMB2016, Poster-48 (2016-4)
- [31] 藤代昂希, 瀬尾茂人, 竹中要一, 古家雅之, 菊田順一, 石井優, 松田秀雄, オプティカルフローを用いた 3 次元生体イメージングデータからの細胞動態の解析, 第 201 回 CVIM 研究会 (2016-3)
- [32] 藏重昂平, 瀬尾茂人, 竹中要一, 松田秀雄, コンポーネントツリーとデータアソシエーションを用いた細胞追跡手法, 第 107 回数理解モデル化と問題解決研究発表会 2016-MPS-107-6 (2016-3)
- [33] 水野雄太, 瀬尾茂人, 渡邊誓旅, 竹中要一, 平松拓郎, 後藤剛, 河田照雄, 松田秀雄, Deep learning を用いた脂肪組織画像における細胞の認識, 第 107 回数理解モデル化と問題解決研究発表会 2016-MPS-107-10 (2016-3)
- [34] 齋藤和正, 瀬尾茂人, 竹中要一, 松田秀雄, 小サンプル数データに対するベイジアンネットワークのスコアベース構造学習改善手法, 第 45 回バイオ情報学研究発表会 2016-B10-45-1 (2016-3)
- [35] 草田義昭, 瀬尾茂人, 竹中要一, 野口眞三郎, 松田秀雄, 乳癌データベースを用いた遺伝子発現プロファイルの数値変換の検討, 第 45 回バイオ情報学研究発表会 2016-B10-45-7 (2016-3)
- [36] Yoichi Takenaka Identification of the state-change-time-point of gene regulations from time-course experiments, The second workshop on Advanced Methodologies for Bayesian Networks

(2015-11)

- [37] Yoichi Takenaka, Shigeto Seno, Hideo Matsuda, Detecting the shifts of gene regulatory networks during time-course experiments with a single time point temporal resolution Proceeding of GIW/InCoB2015, J13 (2015-9)
- [38] 吉田拓真, 瀬尾茂人, 竹中要一, 松田秀雄, 共通 k-mer 種別数に基づくメタゲノム分類手法, 第4回 NGS 現場の会 (2015-7)
- [39] 竹中要一, 風が吹くといつ桶屋は儲かるの? 遺伝子発現制御の時間変化解析, 第4回 NGS 現場の会 (2015-7)
- [40] 松原周平, 瀬尾茂人, 菊田順一, 竹中要一, 石井優, 松田秀雄, オプティカルフローを用いた血管中を遊走する白血球の動態の解析手法, バイオイメージングフォーラム・ワークショップ 2015 (2015-6)
- [41] 藏重昂平, 瀬尾茂人, 間下以大, 菊田順一, 竹中要一, 石井優, 松田秀雄, コンポーネントツリーを用いたグローバルデータアソシエーションによる細胞追跡手法, バイオイメージングフォーラム・ワークショップ 2015, (2015-6)
- [42] 上木怜, 瀬尾茂人, 竹中要一, 松田秀雄, ダイナミックベイジアンネットワークを用いた遺伝子制御ネットワーク推定の部分問題化による近似解法, 情報処理学会研究報告, 第42回バイオ情報学研究会 (2015-6) [図書](計 0件)

〔産業財産権〕

出願状況(計 0件)

名称:
発明者:
権利者:
種類:
番号:
出願年:
国内外の別:

取得状況(計 0件)

名称:
発明者:
権利者:
種類:
番号:
取得年:
国内外の別:

〔その他〕

ホームページ等

6. 研究組織

(1)研究分担者 なし

(2)研究協力者 なし

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。