

平成 22 年 5 月 1 日現在

研究種目： 基盤研究(B)
 研究期間： 2006 ～ 2009
 課題番号： 18340024
 研究課題名(和文) 符号および遺伝子解析実験に共在する組合せ構造とアルゴリズムの研究
 研究課題名(英文) Combinatorial structures and algorithms commonly included in codes
 and pooling designs for genetic experiments

 研究代表者
 神保 雅一 (JIMBO, Masakazu)
 名古屋大学・大学院情報科学研究科・教授
 研究者番号：50103049

研究成果の概要 (和文)：

本研究では、遺伝子解析のための pooling experiment に用いられるグループテストの組合せ構造と positive 識別アルゴリズムが、情報通信における LDPC 符号のパリティ検査行列の組合せ構造や復号アルゴリズムに類似していることに注目し、それらに共通する組合せ構造やアルゴリズムの数理的性質の解明をテーマとして研究を行った。LDPC 符号の復号アルゴリズムと類似の Belief Propagation アルゴリズムを開発し、そのアルゴリズムが pooling experiment の positive 識別アルゴリズムにも有効であることを示した。また、そのための pooling design の持つべき組合せ構造を明らかにした。さらに、これらのグループテストの手法および組合せ構造の情報通信、暗号・セキュリティの分野への応用を試みた。

研究成果の概要 (英文)：

In this research project, we elucidate combinatorial structures and algorithms commonly included in codes and pooling designs for genetic experiments by investigating that the combinatorial structure and algorithm for pooling designs are similar to those of LDPC codes. We developed a belief propagation algorithm to detect positive items among large amount of items and showed that it is effective also for our positive detecting algorithm. And, we clarify the property of combinatorial structures which are useful for our algorithm. Moreover, we consider applications of combinatorial designs to the fields of codes, information transmission and cryptography.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2006 年度	4,000,000	1,200,000	5,200,000
2007 年度	3,500,000	1,050,000	4,550,000
2008 年度	2,600,000	780,000	3,380,000
2009 年度	2,700,000	810,000	3,510,000
年度			
総計	12,800,000	3,840,000	16,640,000

研究分野：数物系科学

科研費の分科・細目：数学・数学一般(含確率論・統計数学)

キーワード：pooling experiment, グループテスト, positive detecting algorithm, 確率伝搬法, BP, CCCP, LDPC

1. 研究開始当初の背景

研究代表者は、組合せデザインなどの離散数学およびその応用に興味を持って研究を行ってきたが、近年符号理論において LDPC (Low Density Parity Check) 符号と呼ばれる疎なパリティ検査行列を持つ符号のクラスの復号アルゴリズムに sum-product アルゴリズムと呼ばれる BP(Belief Propagation) アルゴリズムの一種が有用であることが示された。一方で、グループテストと呼ばれる分野では、多くの item が含まれている集団の中に、ある特徴を持つ少数の item が含まれているときに、それらの少数の item を効率よく識別するために、個別に 1 つ 1 つの item をテストするのではなく、様々な組合せの item の部分集合(グループと呼ぶ)を多数作って、各グループに対してテストを行い、その結果から、positive item を識別する検査法が用いられる。従来、その識別アルゴリズムには、組合せ論的手法が用いられていたが、研究代表者は、グループテストの分野にも sum-product アルゴリズムと同様の belief propagation アルゴリズムが有効であることを予想してアルゴリズムの開発に着手した。また、そのための効率的な組合せ構造の特徴を見出すことも重要なテーマとして研究を開始した。

2. 研究の目的

pooling experiment では、A,T,G,C の塩基からなる DNA の多数の断片にある試験を施して陽性(positive)反応を示す clone を見出す試験がよく行なわれる。このテストの際に、複数の塩基列をひとつにまとめていろいろなグループ(pool と呼ばれる)を作り、それらに対して反応試験を行ない、その実験結果から positive clone を見出すことにより実験回数を大幅に削減できる。この方法をグループテストと読んでいます。しかし、このような pooling experiment において、各実験の結果に false negative, false positive などの誤りが生じることは避けられない。目的の clone を効率よく見出すためには、結果に誤りがあることを想定して、誤り訂正能力を有する pooling design を計画しなければならない。ここに、符号理論の誤り訂正の概念が重要となる。一方、得られた実験結果から各 clone がポジティブである確率を高速かつ効率的に求めるためのポジティブ識別アルゴリズムが必要である。このアルゴリズムにベイジアンネットワークのアルゴリズムが有効であることを示すことが本研究の第 1 の目的である。このアルゴリズムと類似のアルゴリズムが LDPC 符号にも用いられており、パリティ検査行列を隣接行列とする二部グラフに短いサイクルが存在するとベイジアンネ

ットワークによる復号アルゴリズムが収束しないあるいは収束値の精度が低くなるということが知られている。したがって、短いサイクルを持たないパリティ検査行列の組合せ構造の研究が重要である。本研究では、グラフのサイクルの長さや構造に注目し、光直交符号、量子符号などとの関係も視野に入れて、符号、pooling design に共在する組合せ構造とアルゴリズムの収束性の関係を明らかにすることを第 2 の目的とした。

さらに、情報通信における組合せ構造の有用性にも注目し、最適な光直交符号、conflict avoiding code, Frequency hopping 系列などの組合せ論的構成法の研究も併せて行うことを第 3 の目的とした。

3. 研究の方法

研究は下記のように実施した。

(1) pooling design の組合せ構造の研究に焦点を絞り、検査結果に誤りが含まれている場合に、誤りがある数以下であれば、正確に positive なクローンを判別できる pooling design を構成する方法を組合せ論的手法を用いて見出した。特に、pooling design の結合関係を表す 2 部グラフに長さ 4 のサイクルが無い場合に、最適な pooling design の存在・構成問題を研究した。

(2) また、clone の全順序が分かっており、連続する高々 2 つ (あるいは複数個) の clone が positive であるという場合には、誤り訂正能力を有する pooling design の構成法が有効であり、 $d=3$ あるいはそれ以上の最小距離を持つ最適な design の構成法に関する研究を行った。

(3) 確率伝播法を用いた新しいポジティブ識別アルゴリズムを開発し、その効率・収束性について、研究を行った。平成 18 年度はまず、CCCP (Convex-Concave Procedure) と呼ばれる手法を基本としたアルゴリズムを構築し、長さ 4 のサイクルがある場合にも適用可能かどうかを調べ、さらに、CCCP 以外の独自のアルゴリズムの開発の可能性を探った。また、シミュレーションを行い、その効率を評価した。さらに、従来の MCMC による方法、および、本研究の研究代表者らによって開発されたベイジアンネットワークによる方法との比較を行った。

(4) 光直交符号、conflict avoiding code などの新しい符号を組合せデザインを用いて構成する方法の研究を行う。また、これらの符号の復号アルゴリズムについても(ii)の分担者

と協力をしながら研究を行ってきた。

(5) 研究代表者・神保は、分担者と共に、定期的に会合を行い、研究連絡を密にして上記の研究を遂行してきた。その際、研究分担者・水島の協力を得て、バイオインフォーマティクスの立場から実際の実験への適用可能性も評価を行ってきた。

4. 研究成果

平成 18 年度から平成 21 年度までの 4 年間に下記の研究成果を得た。

(1) pooling experiment において、clone と pool(group)の結合関係を表す Tanner graph に長さ 4 のサイクルが存在しないときには、positive 識別アルゴリズムに belief propagation(BP)のアルゴリズムを用いることにより効率よく positive item を識別することができることを示した。この結果は、bioinformatics の分野で評価の高い IEEE/ACM bioinformatics に掲載された。

(2) 一方、Tanner graph に長さ 4 のサイクルが多数存在するときには BP によるアルゴリズムは収束しないことが多く、収束してもその収束値の真の値からの偏りが大きくなることをシミュレーションにより示し、この場合には、BP ではなく CCCP (Convex-Concave Procedure) によるアルゴリズムと MCMC(Markov Chain Monte Carlo)による hybrid algorithm が有効であることを示した。この結果は統計学会誌に掲載された。

(3) (1), (2)のアルゴリズムの有効性はシミュレーションにより実証したものであるが、情報幾何的手法により、Tanner graph に長さ 4 のサイクルが存在しないときには、その収束値の最大剰余項がゼロになり、近似の精度が良いことを理論的に示した。その結果は、現在、論文としてまとめ、投稿中である。

(4) さらに、グループテストによる fingerprint 符号の結託者発見のためのアルゴリズムに BP を用いることを提案し、その効率を調べた。

(5) 最適な pooling design の組合せ論的特徴は d-disjunct あるいは unique colinearity condition などの組合せデザインのもつ性質と関連が深いため、そのようなデザインの構成法を見出し、光直交符号、conflict-avoiding code, frequency hopping sequence などとの関連についてもいくつかの研究成果を得た。これらの結果は、SIAM Discrete Math.,

IEEE/IT などの当該分野のトップジャーナルに掲載された。

(6) 光直交符号などに有用な Steiner quadruple system(SQS)の構成法について研究を継続し、すべての単元を multiplier として持つ strictly cyclic SQS の 2 種類の逐次構成法を得た。また、計算機を用いて、パラメータが小さい場合に所望の性質をもつ SQS を得た。これらにより、すべての単元を multiplier として持つ strictly cyclic SQS についての多くの無限系列を得た。本結果は、論文としてまとめて、論文誌に投稿予定である。また、2009 年 11 月に台湾の国立交通大学を訪問した際に、本研究内容について講演を行った。

本研究における研究成果はおおむね当初の予定通りの成果を得られたと考えている。(1)から(3)について、当初、pooling experiment のためのアルゴリズムを開発し、その良さを評価するという計画を立てていたが、このために、BP によるアルゴリズムを開発し、その効率をシミュレーションにより評価し、Tanner graph に長さ 4 のサイクルがないときにはその収束の良さを理論的に示すことができた。このアルゴリズムは、遺伝子情報解析の分野にとどまらず、spam mail の識別など、広くグループテストの分野に応用できる一般的なアルゴリズムとして、応用範囲が広いと思われる。

光直交符号や conflict-avoiding code, frequency hopping sequence への組合せデザインへの応用についても十分な成果が得られている。また、fingerprint code への新たな応用を見出したことは予想以上の成果であると言って良いであろう。

今後の課題：

今後、Tanner graph に長さ 4 のサイクルが存在するときに、BP, CCCP などの収束値の真の値からの偏りを理論的に評価することが課題である。また、Fingerprint code への応用については、その Tanner graph が pooling experiment や LDPC 符号と異なり、item の数が group の数より少ないため、収束の様子が異なる。この理論的解析も課題として残っている。また、開発したアルゴリズムの新たな応用分野を見出し、実践的な研究へと展開することも今後の課題であろう。

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 21 件)

- [1] H. Uehara and M. Jimbo, A positive detecting algorithm for DNA library screening based on CCCP. J. Japan Statist. Soc., Vol. 39 (2009) 89-109. 査読有
- [2] H. Uehara and M. Jimbo, A Positive Detecting Code and its Decoding Algorithm for DNA Library Screening. IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol. 6 (2009) 652-666. 査読有
- [3] C. Ding, R. Fuji-Hara, Y. Fujiwara, M. Jimbo and M. Mishima, Sets of frequency hopping sequences: bounds and optimal constructions, to appear in IEEE Transactions on Information Theory. Vol. 55 (2009) 3297-3304. 査読有
- [4] M. Jimbo, K. Shiromoto, A construction of mutually disjoint Steiner systems from isomorphic Golay codes. J. Combinatorial Theory, Ser. A, Vol. 116 (2009) 1245-1251. 査読有
- [5] K. Ozawa, N. Tsushima, M. Jimbo, A-optimal diarell cross experiments for estimating g.c.a. effects, Statistics and Application, Vol. 4 (2009) 印刷中, 査読有
- [6] M. Mishima, H.-L. Fu and S. Uruno, Optimal conflict-avoiding codes of length $n \equiv 0 \pmod{16}$ and weight 3, to appear in Designs, Codes and Cryptography, Vol.52 (2009) 275-291. 査読有
- [7] C.-C. Chou, C.-M. Fu, T. Minoura and M. Mishima, Cycle Decomposition of 2-fold complete tripartite graphs and generalized pseudo-characteristic, to appear in Journal of Statistics and Applications. Vol. 4 (2009) 397-408. 査読有
- [8] I. Mejza, S. Mejza and S. Kuriki, Square lattice designs in incomplete split-block designs, Journal of Statistics and Applications, Vol. 4 (2009) 327-338. 査読有
- [9] S. Mejza, S. Kuriki and D. Kachlicka, Repeated Youden Squares with subplot treatments in a group-divisible design, Journal of Statistics and Applications, Vol. 4 (2009) 369-377. 査読有
- [10] S. Mejza, S. Kuriki and D. Kachlicka, Repeated Youden Squares with subplot treatments in a proper incomplete block design, Biometrical Letters, Vol. 46 (2009) 153-162. 査読有
- [11] S. Yagi, K. Mimura, M. Jimbo, A construction of $OA(s^t, t+1, s, t)$ by nonlinear functions and some classification for $s=4$. Journal of Statistical Planning and Inference. Vol. 138, 3309-3315, 2008. 査読有
- [12] M. Sawa, S. Kageyama and M. Jimbo, Compatibility of BIB designs. Statistics and Applications, Vol. 6 (2008) 56-71. 査読有

- [13] T. Hishida, M. Jimbo, M. Mishima, K. Ozawa, Y. Mutoh, Further constructions for BIB designs with nested rows and columns. Ars Combinatoria, Vol. 86 (2008) 239-256. 査読有
- [14] K. Momihara, M. Jimbo, Some constructions for block sequences for Steiner quadruple systems with error correcting consecutive unions. Journal of Combinatorial Designs, Vol.16 (2008), 152-163. 査読有
- [15] M. Mishima, The spectrum of 1-rotational Steiner triple systems over a dicyclic group, Discrete Mathematics vol.308, no.12 (2008), 2617-2619. 査読有
- [16] M. Ohashi, N. Sakamoto, M. Watanabe, H. Mizushima, H. Tanaka, Development of a Telediagnosis endoscopy system over secure Internet. Methods Inf. Med., Vol. 47 (2008) 157-166. 査読有
- [17] M. Jimbo, M. Mishima, A.Y. Janiszewski, V.D. Teymorian, V. Tonchev, On conflict-avoiding codes of length $n=4m$ for three active users. IEEE Transactions on Information Theory, Vol. 53 (2007) 2732-2742. 査読有
- [18] K. Momihara, M. Mueller, J. Satoh and M. Jimbo, Constant weight conflict-avoiding codes. SIAM Journal on Discrete Mathematics, Vol. 21 (2007) 959-979. 査読有
- [19] K. Ozawa, M. Mishima, S. Kuriki, M. Jimbo, Constructions for rectangular designs. Utilitas Mathematicae, Vol. 71 (2007) 179-196. 査読有
- [20] M. Mueller, M. Jimbo, Cyclic sequences of k -subsets with distinct consecutive unions. Discrete Mathematics, 308 (2006) 457-464. 査読有
- [21] K. Ozawa and S. Kuriki, Incomplete split-plot designs generated from α -resolvable designs. Statistics & Probability Letters, Vol. 76 (2006) 1245-1254. 査読有

〔学会発表〕 (計 21 件)

- [1] 田口和規, 栗木進二, 分割型ユニットをもつ 2 因子実験の構成法, 日本数学会 2010 年度年会, 2010 年 3 月 25 日, 慶応義塾大学
- [2] 国原雄太, 澤正憲, 神保雅一, 種々の性質をもつ 3-デザインの構成法とその応用, 2009 年度応用数学合同研究集会, 2009 年 12 月, 滋賀.
- [3] 神保雅一, On a strictly cyclic SQSs admitting all units as its multipliers. 研究集会「組合せデザイン理論とその応用」, 2009 年 8 月 31 日, 湯河原.
- [4] S. Kuriki, A construction method of incomplete split-plot designs supplemented by control treatments. 4th Workshop on Statistics, Mathematics and Computation, 2009 年 11 月 9

- 日, Aberta 大学 (ポルトガル) .
- [5] 田口和規, 栗木進二, 分割型ユニットをもつ2因子実験の構成法, 研究集会「組合せデザイン理論とその応用」, 2009年8月31日, 湯河原.
- [6] S. Mase, Y. Murayama, On a simulation of Gaussian random fields. Workshop on Models and Iages of Porous Media, 2009年1月15日, Universite Paris Descartes.
- [7] M. Jimbo, K. Momihara, S. Yoshikawa, A strictly cyclic Steiner quadruple system on Z_v admitting all units as multipliers. The 4th International Conference on Combinatorial Mathematics and Combinatorial Computing, 2008年12月16日, ニューージーランド, オークランド大学
- [8] 間瀬茂, 山根翔太, Prediction of surface distributions of seismic intensities using a hierarchical Bayes model. 科研費シンポジウム「時空間現象データに対する統計科学モデルの構築および解析に関する組織的研究」, 2008年11月28日, 沖縄県青年会館.
- [9] M. Jimbo, A Steiner quadruple systems on Z_p^m admitting all units as multipliers. Combinatorial Design Theory, 2008年11月13日, カナダ, Banff International Research Station招待講演
- [10] 神保雅一, 城本啓介, Quantum jump codeと組合せデザイン, 代数的符号理論と組合せデザイン, 2008年10月15日, 京都大学, 数理解析研.
- [11] H. Uehara, M. Jimbo, Group testing algorithm for DNA library screening based on BP and CCCP. Group testing in life science, 2008年7月13日, ドイツ, Dagstuhl, 招待講演.
- [12] 三嶋美和子, 神保雅一, Cyclic Steiner Quadruple System の再帰的構成法について, 研究集会「実験計画法と統計的推測理論の展開」, 2007年11月28日, 城崎.
- [13] 柴田雄規, 神保雅一, Random fingerprinting codeに対する確率的解析, 研究集会「実験計画法と統計的推測理論の展開」, 2007年11月27日, 城崎
- [14] 神保雅一, Quantum jump codeと組合せデザイン, 「計算代数統計学の展開」科研費研究集会, 2007年10月26日, 豊橋.
- [15] M. Jimbo, Quantum jump codes derived from affine geometry and cyclic codes. 2007 COE Conf. Dev. Dyn. Math.High Func., Oct. 2, 2007, Fukuoka.
- [16] 神保雅一, 組合せデザインとその情報通信への応用, 組合せ論サマースクール, 2007年9月3日, 沖縄宜野湾市.
- [17] K. Momihara, M. Mueller, J. Satoh, M. Jimbo, Bounds and constructions for optimal

- constant weight conflict-avoiding codes. 2007 IEEE International Symp. on Information Theory, 2007年6月25日, Nice, France.
- [18] M. Mishima, On conflict-avoiding codes of length $n=4m$ for three active users. International Workshop on Combinatorics 2007, 2007年6月12日, 京都大学数理解析研究所.
- [19] 神保雅一, 上原啓明, ゲノム塩基配列識別アルゴリズムとベイジアンネットワーク, 応用統計学会, 2007年5月24日, 東工大, 特別講演

〔図書〕 (計4件)

- [1] S. Mase 他, M. Lovric (ed.), International Encyclopedia of Statistical Science, (Geostatistics の解説担当), 2010年発行予定
- [2] S. Mase, Bayesian Networks A Practical Guide to Applications: Chapter 15 Credit-rating of companies, 2008, 263-277, Wiley.
- [3] 間瀬茂, R プログラミングマニュアル, 数理工学社, 2007, 335頁.
- [4] 長尾壽夫, 栗木進二, 数理解析学, 共立出版, 2006, 250頁.

6. 研究組織

(1) 研究代表者

神保 雅一 (JIMBO MASAKAZU)
名古屋大学・大学院情報科学研究科・教授
研究者番号: 50103049

(2) 研究分担者

間瀬 茂 (MASE SHIGERU)
東京工業大学・情報理工学研究所・教授
研究者番号: 70108190

松本 裕行 (MATSUMOTO HIROYUKI)
名古屋大学・大学院情報科学研究科・教授
研究者番号: 00190583

栗木 進二 (KURIKI SHINJI)
大阪府立大学・工学研究科・教授
研究者番号: 00167389

三嶋美和子 (MISHIMA MIWAKO)
岐阜大学・工学部・准教授
研究者番号: 00283284

水島 洋 (MIZUSHIMA HIROSHI)
東京医科歯科大学・情報処理センター・特任准教授
研究者番号: 50219630

金森 敬文 (KANAMORI TAKAFUMI)
名古屋大学・情報科学研究科・准教授
研究者番号: 60334546
(H19~H21年度: 研究分担者)

(3) 連携研究者

太田 克弘 (OTA KATSUHIRO)

慶應義塾大学・理工学部・教授

研究者番号：40213722

(H18～H19：研究分担者，H20～21：連携研究者)

菱田 隆彰 (HISHIDA TAKAAKI)

愛知工業大学・工学部・准教授

研究者番号：30329627

(H18～H19：研究分担者，H20～21：連携研究者)

小澤 和弘 (OZAWA KAZUHIRO)

岐阜県立看護大学・看護学部・准教授

研究者番号：20336629

(H18～H19：研究分担者，H20～21：連携研究者)