

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存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 inherited both in codes and pooling designs for genetic analysis

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

研究分野：数物系科学
科研費の分科・細目：数学・数学一般
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. 研究計画の概要

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を解明することを目的としている。そのために、これらのいずれにも有効な組合せ構造の特徴を見出し、その特徴を有する組合せデザインを構成し、同時に LDPC 符号の復号アルゴリズムと類似のアルゴリズムが pooling experiment の positive 識別アルゴリズムにも用いられることを示す。

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を情報通信、暗号・セキュリティの分野にも応用することを試みる。

2. 研究の進捗状況

平成18年度から平成20年度までの3年間に下記の研究成果を得ている。

(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、IEEE/ACM bioinformatics に掲載予定である。

(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予定である。

(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、論文としてまとめている。

(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こと、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。

3. 現在までの達成度

②おおむね順調に進展している。

(理由)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、この計画は、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。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。

4. 今後の研究の推進方策

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として残っている。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。最後の 1 年はこれらの課題に取り組むとともに, 他の情報通信分野への応用についても研究を展開してゆきたい。

5. 代表的な研究成果

[雑誌論文] (計 10 件)

- [1] H. Uehara and M. Jimbo, A positive detecting algorithm for DNA library screening based on CCCP. J. Japan Statist. Soc. 2009 印刷中, 査読有
- [2] 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. 2009 印刷中, 査読有
- [3] Mejza, I., Mejza, S. and Kuriki, S. (2008) Square lattice designs in incomplete split-block designs, Journal of Statistics and Applications 2009 印刷中, 査読有
- [4] Mejza, S., Kuriki, S. and Kachlicka, D. (2009) Repeated Youden Squares with subplot treatments in a group-divisible design, Journal of Statistics and Applications 2009 印刷中, 査読有
- [5] 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. (DOI: 0.1007/s10623-009-9282-2). 2009 印刷中, 査読有
- [6] 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. 2009 印刷中, 査読有
- [7] 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. 5, No. 6, DOI: 10.1109/TCBB.2007.70266, 2008. 査読有

[8] 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

[9] 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

[10] 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

[学会発表] (計 7 件)

[1] 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, ニュージーランド、オークランド大学

[2] 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

[3] 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.

[4] 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.

[5] 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

[6] 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.

[7] 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