

様式 C-7-1

平成19年度科学研究費補助金実績報告書（研究実績報告書）

1. 機関番号 

|   |   |   |   |   |
|---|---|---|---|---|
| 1 | 4 | 6 | 0 | 3 |
|---|---|---|---|---|

 学 2. 研究機関名 奈良先端科学技術大学院大
3. 研究種目名 若手研究 (B) 4. 研究期間 平成19年度～平成20年度
5. 課題番号 

|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| 1 | 9 | 7 | 1 | 0 | 1 | 7 | 2 |
|---|---|---|---|---|---|---|---|
6. 研究課題名 高次元オミックスデータとヘテロ医療診断データの統合的解析法の開発

7. 研究代表者

| 研究者番号           | 研究代表者名 | 所属部局名   | 職名 |
|-----------------|--------|---------|----|
| 8 0 3 6 2 8 3 8 | 大羽, 成征 | 情報科学研究科 | 助教 |

8. 研究分担者(所属研究機関名については、研究代表者の所属研究機関と異なる場合のみ記入すること。)

| 研究者番号 | 研究分担者名 | 所属研究機関名・部局名 | 職名 |
|-------|--------|-------------|----|
|       | カガナ    |             |    |
|       | カガナ    |             |    |
|       | カガナ    |             |    |
|       | カガナ    |             |    |
|       | カガナ    |             |    |

9. 研究実績の概要(国立情報学研究所でデータベース化するため、600字～800字で記入。図、グラフ等は記載しないこと。)

下欄には、当該年度に実施した研究の成果について、その具体的内容、意義、重要性等を、交付申請書に記載した「研究の目的」、「研究実施計画」に照らし、600字～800字で、できるだけ分かりやすく記述すること。また、国立情報学研究所でデータベース化するため、図、グラフ等は記載しないこと。

近年の臨床医学データ解析では臨床診断情報に加えて遺伝子発現量やSNPなどの高次元変量オミックスデータを同時に扱う必要があるため(1)各症例の特徴を示す変量の高次元性・ヘテロ性・欠測などの問題、(2)解析結果出力時の統計的検定多重性の問題への対応が難しい。本研究課題ではこれらを統一的に解決する方法の開発と普及を目指している。私はまず欠測を含む行列形データの因子化を柔軟な確率モデルで記述することで(1)の問題を扱うための基本的プラットフォームとした(これは本研究の計画段階で中心においていた確率的正準相関解析(CCA)の一般化という位置づけ)。とくに欠測の構造にヘテロ性が入っている場合にスパースな因子負荷行列を求める「確率的ヘテロ成分分析」(NIPSスポットライトトークに選ばれる)、各成分と各症例の信頼性の違いが同時に存在する場合を考慮して重み付けを工夫した因子分解法「縦横二方向因子分析」(IBIS, AROBなど)、データがデジタル値である場合に同様の因子分解を行う「重みつきマージン最大化行列分解」(AROB)を開発し発表した。(2)については、多重検定における統計的有意性スコア最適化原理(ODP)の理論と応用に関する研究を進めた。とくにODPの推定法を改善し検出力を向上する工夫(NC研究会)および、二値共変量との間に有意な関係を持つ遺伝子の検出問題において、関係のパターンに関する情報を取り入れることで検出力を高めたアイディア(IBIS)を発表した。高次元変量データを疾病診断などの症例分類につなげる方法論の研究には一段落がつきつつあるが、疾病のメカニズムに迫るためのデータ解析の手法にはまだ大きな進歩の余地がある。次年度では(1)と(2)の方向性を統合する手法開発と、その普及を目指す予定である。

※ 成果の公表を見合わせる必要がある場合は、その理由及び差し控え期間等を記入した調書(A4判縦長横書1枚)を添付すること。

10. キーワード

- |               |           |          |
|---------------|-----------|----------|
| (1) 遺伝子発現量データ | (2) 多変量解析 | (3) 多重検定 |
| (4) 行列因子化     | (5)       | (6)      |
| (7)           | (8)       |          |

(裏面に続く)

## 11.研究発表（平成19年度の研究成果）

〔雑誌論文〕 計（ 0 ）件

| 著者名 | 論文標題  |   |       |         |
|-----|-------|---|-------|---------|
|     |       |   |       |         |
| 雑誌名 | 査読の有無 | 巻 | 発行年   | 最初と最後の頁 |
|     |       |   | ・ ・ ・ |         |

| 著者名 | 論文標題  |   |       |         |
|-----|-------|---|-------|---------|
|     |       |   |       |         |
| 雑誌名 | 査読の有無 | 巻 | 発行年   | 最初と最後の頁 |
|     |       |   | ・ ・ ・ |         |

| 著者名 | 論文標題  |   |       |         |
|-----|-------|---|-------|---------|
|     |       |   |       |         |
| 雑誌名 | 査読の有無 | 巻 | 発行年   | 最初と最後の頁 |
|     |       |   | ・ ・ ・ |         |

〔学会発表〕 計（ 10 ）件

| 発表者名  | 発表標題                              |             |  |
|---|-----------------------------------|-------------|--|
| 大羽 成征, 川鍋 一晃, K.R.Mueller, 石井 信   | Heterogeneous Component Analysis. |             |  |
| 学会等名  | 発表年月日                             | 発表場所        |  |
| 21st annual conference on Neural Information Processing Systems (NIPS 2007) | 2007年12月3日-5日                     | カナダ, バンクーバー |  |

| 発表者名  | 発表標題   |      |  |
|---|--|------|--|
| 袖林 和宏, 大羽 成征, 石井 信  | Two-Way Factor Analysis for Missing Value Estimation of Matrix Data. |      |  |
| 学会等名  | 発表年月日  | 発表場所 |  |
| International Symposium on Artificial Life and Robotics (AROB 13th '08) | 2008年1月31日-2月2日  | 別府   |  |

| 発表者名  | 発表標題   |      |  |
|---|--|------|--|
| 野村 啓仁, 大羽 成征, 石井 信  | Re-weighted ODP for differential gene expression analysis. |      |  |
| 学会等名  | 発表年月日  | 発表場所 |  |
| International Symposium on Artificial Life and Robotics (AROB 13th '08) | 2008年1月31日-2月2日  | 別府   |  |

| 発表者名  | 発表標題   |           |  |
|---|--|-----------|--|
| 古谷 允宏, 大羽 成征, 石井 信  | Collaborative filtering based on a Weighted Maximum Margin Matrix Factorization. |           |  |
| 学会等名  | 発表年月日  | 発表場所      |  |
| International Symposium on Artificial Life and Robotics (AROB 13th '08) | 2008年1月31日-2月2日  | 別府Bconプラザ |  |

| 発表者名               | 発表標題                 |      |  |
|--------------------|----------------------|------|--|
| 野村 啓仁, 大羽 成征, 石井 信 | 多重検定におけるODP統計量の有効な推定 |      |  |
| 学会等名               | 発表年月日                | 発表場所 |  |
| ニューロコンピューティング研究会   | 2008年3月12日-14日       | 玉川大学 |  |

|                           |                      |                     |
|---------------------------|----------------------|---------------------|
| 発表者名                      | 発表標題                 |                     |
| 袖林 和宏, 大羽 成征, 石井 信        | 二方向因子分析による行列データの欠損予測 |                     |
| 学会等名                      | 発表年月日                | 発表場所                |
| 情報論的学習理論ワークショップ(IBIS2007) | 2007年11月5日-7日        | 東京工業大学大学 すすかけ台キャンパス |

|                           |                        |                     |
|---------------------------|------------------------|---------------------|
| 発表者名                      | 発表標題                   |                     |
| 大羽 成征, 石井 信               | 隠れ変数モデルに基づく同時検定用統計量最適化 |                     |
| 学会等名                      | 発表年月日                  | 発表場所                |
| 情報論的学習理論ワークショップ(IBIS2007) | 2007年11月5日-7日          | 東京工業大学大学 すすかけ台キャンパス |

|             |   |      |
|-------------|---|------|
| 発表者名        | 発表標題  |      |
| 大羽 成征, 石井 信 | 多重検定のための検定統計量最適化：隠れ変数を仲立ちとした情報共有に基づく新しいアプローチについて. |      |
| 学会等名        | 発表年月日   | 発表場所 |
| 統計関連学会連合大会  | 2007年9月6日-9日                                      | 神戸大学 |

|                    |                    |      |
|--------------------|--------------------|------|
| 発表者名               | 発表標題               |      |
| 袖林 和宏, 大羽 成征, 石井 信 | 二方向因子分析による欠測データ予測. |      |
| 学会等名               | 発表年月日              | 発表場所 |
| 統計関連学会連合大会         | 2007年9月6日-9日       | 神戸大学 |

|                            |                              |                        |
|----------------------------|------------------------------|------------------------|
| 発表者名                       | 発表標題                         |                        |
| 大羽 成征, 富岡 伸元, 大平 美紀, 石井 信. | アレイ比較ゲノムデータ正規化手法Combfitについて. |                        |
| 学会等名                       | 発表年月日                        | 発表場所                   |
| バイオ情報学研究会                  | 2007年6月13日-14日               | 沖縄科学技術研究基盤整備機構シーサイドハウス |

〔図書〕 計 ( 0 ) 件

|     |     |     |       |
|-----|-----|-----|-------|
| 著者名 | 出版社 |     |       |
|     |     |     |       |
|     | 書名  | 発行年 | 総ページ数 |
|     |     |     |       |

12. 研究成果による産業財産権の出願・取得状況

〔出願〕 計 ( 0 ) 件

| 産業財産権の名称 | 発明者 | 権利者 | 産業財産権の種類、番号 | 出願年月日 | 国内・外国の別 |
|----------|-----|-----|-------------|-------|---------|
|          |     |     |             |       |         |

〔取得〕 計 ( 0 ) 件

| 産業財産権の名称 | 発明者 | 権利者 | 産業財産権の種類、番号 | 取得年月日 | 国内・外国の別 |
|----------|-----|-----|-------------|-------|---------|
|          |     |     |             |       |         |

13. 備考

※ 研究者又は所属研究機関が作成した研究内容又は研究成果に関するwebページがある場合は、URLを記載すること。

<http://hawaii.naist.jp/~shige-o/>