

平成 21 年 6 月 26 日現在

研究種目：基盤研究（C）
研究期間：2007～2008
課題番号：19500250
研究課題名（和文） 蛋白発現データ解析のための統計モデルの研究と癌個別診断の試み
研究課題名（英文） statistical model for protein expression data analysis and its application to individualized medicine for cancer patients
研究代表者
宮田 敏（MIYATA SATOSHI）
財団法人 癌研究会 ゲノムセンター 研究員
研究者番号：60360343

研究成果の概要：

質量分析計からの二次元蛋白データは、相関構造を持った TOF/MS データの集合であり、蛋白発現プロファイルの関数データの抽出には inhomogeneously smooth な関数の平滑化が必要である。このため、不連続点あるいは微分不能な点を含む目的関数を、ノイズを含んだデータから推定するのに適した adaptive free-knot spline の理論的研究およびプログラム開発を行った。また、蛋白発現データを微分不能な点を持つ確率分布から得られたものと解釈し、きわめて特異な確率密度関数の推定問題として研究した。さらにこれらを用いた判別問題への応用も検討した。

これら不均一な滑らかさを持つ関数データの抽出、確率密度関数の推定、判別問題の解決は、多重節点を持つスプライン関数を用いてモデル化された。スプライン関数の tuning parameter である節点の最適化は、適応型モデル選択基準（adaptive model selection criterion）を最適化することで行われた。この最適化問題は複雑な非線形問題となるが、遺伝的アルゴリズムを適用することで、大域的な最適解を探索することで解決した。

研究の成果は、日本癌学会学術総会、統計関連学会連合大会などの国内学会、および、台北大学で行われた中国統計学会、日本統計学会、韓国統計学会合同の joint statistical conference of CSA-JSS-KSS での招待講演で報告した。また主要研究成果は論文発表のため投稿準備中である。

交付額

（金額単位：円）

	直接経費	間接経費	合計
19年度	1,900,000	570,000	2,470,000
20年度	1,500,000	450,000	1,950,000
年度			
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：総合領域

科研費の分科・細目：情報学・生態生命情報学

キーワード：プロテオーム情報処理

1. 研究開始当初の背景

癌において特異的に変化する蛋白質の探

索は、癌の個別化診断や治療方法、抗がん剤の選択、治療効果判定、予後予測等につながる

るものとして期待されている。近年では、組織切片を細かい格子状に分割し grid ごとにレーザーを照射することで、切片上の二次元的な蛋白発現分布を計測する手法も開発されている。

本研究では質量分析計 (TOF/MS, time-of-flight mass spectrometry) によって測定された蛋白質のプロファイリングパターン認識と、治療方法の選択や予後判定の指標となる癌部蛋白質の同定、およびその判別方法の開発を目的とする。

2. 研究の目的

二次元蛋白発現データは新しい型のデータであり定まった解析法は存在しないが、本研究では個々の TOF/MS データの蛋白同定に開発済みの手法を組織切片上の grid ごとに適用した上で、関数データ解析、空間統計学の手法を合わせた独創的な手法で解析する予定である。

さらにこうして得られた蛋白発現データに遺伝子発現および SNP データを加え、AdaBoost 等の機械学習理論に基づく手法により臨床情報の予測を行うがん患者に対する個別改良の確率に寄与することを目的とする。

3. 研究の方法

研究に用いるサンプルについては、すでに臨床情報の収集 (インフォームドコンセント取得済み)、組織検体および抹消血からの蛋白発現データの取得 (質量分析計 surface enhanced laser desorption and ionization time-of-flight mass spectrometry (SELDI TOF/MS) を行い、血漿蛋白 337 症例、癌組織蛋白 449 症例を解析済み) が完了した。

質量分析計からの二次元蛋白データは、相関構造を持った TOF/MS データの集合であり、蛋白発現プロファイルの関数データの抽出には inhomogeneously smooth な関数の平滑化が必要である。このため、必要な adaptive free-knot spline の理論的研究およびプログラム開発を行った。また、蛋白発現データを微分不能な点を持つ確率分布から得られたものと解釈し、きわめて特異な確率密度関数の推定問題として研究した。さらにこれらを用いた判別問題への応用も検討した。

4. 研究成果

1) 蛋白プロファイルからの peak 同定のため、adaptive free-knot spline を用いたデータの平滑化の後、局所分散との S/N 比を用いた手法を開発、2) 検体間での peak の補正には、各検体における peak 近傍の情報を基に、集団での近隣ピークを分離する手法を開発、3) 臨床情報と関連する蛋白の同定および予測

モデルの構築には、AdaBoost などの機械学習理論に基づく方法を検討した。

Adaptive free-knot spline による蛋白発現データの平滑化には、knot sequence の最適化が必要であるが、そのために適応型モデル選択基準 (adaptive model selection criterion) によるモデルのパフォーマンスの評価を定式化し、その最適化に遺伝的アルゴリズムによる大域的最適解の探索を提案した。

また、蛋白発現データを微分不能な点を持つ確率分布から得られたものと解釈し、不連続点を含む極めて特異な確率分布のノンパラメトリックな推定問題として定式化した。その推定のために、adaptive free-bin histogram と adaptive free-knot logspline 推定量を提案した。研究の成果は論文発表の準備中であるが、確率密度推定の方法論として、不連続点あるいは微分不能な点が存在し、その位置と個数が未知であるというモデルは、世界的に見ても前例のないものである。

研究の成果は論文として発表準備を進めるとともに、日本統計学会、科学研究費研究集会、Joint Session of Chinese Statistical Association (CSA), Japanese Statistical Society (JSS) and Korean Statistical Society (KSS) などで発表した。

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 11 件)

1. Jin, G., Yamazaki, Y., Takuwa, M., Takahara, T., Kaneko, K., Kuwata, T., Miyata, S. and Nakamura, T. (2007), "Trib1 and Evi1 cooperate with Hoxa and Meis1 in myeloid leukemogenesis", *Blood*, Vol. 109, No. 9, 3998-4005. 査読あり
2. Oishi, Y., Nagasaki, K., Miyata, S., Matsuura, M., Nishimura, S., Iwase, T., Kasumi, F., Futoshi Akiyama, F., Sakamoto, G., Iwai, T., and Miki, Y. (2007), "Gene expression analysis of primary lesion of supraclavicular lymph node metastasis-positive breast cancer", *Journal of Human Genetics*, Vol. 52, No. 3, 271-279. 査読あり
3. Fujisawa, H., Isomura, M., Eguchi, S., Ushijima, M., Miyata, S., Miki, Y., and Matsuura, M. (2007), "Identifying haplotype block structure by using ancestor-derived model and MDL principle", *Journal of Human Genetics*, Vol. 52, No. 9, 738-746. 査読あり
4. Ushijima, M., Miyata, S., Eguchi, S.,

- Kawakita, M., Yoshimoto, M., Iwase, T., Akiyama, F., Sakamoto, G., Nagasaki, K., Miki, Y., Noda, T., Hoshikawa, Y. and Matsuura, M. (2007), "Common peak approach by semi-supervised learning using mass spectrometry datasets for predicting effects on anticancer drugs on breast cancer", *Cancer Informatics*, Vol. 3, 285-293. 査読あり
5. Enomoto, Y., Ushijima, M., Miyata, S., Matsuura, M., Ohtaki, M. (2008), "Inference on biological mechanisms using an integrated phenotype prediction model", *Hiroshima J Med Sci.*, Vol. 57, No. 1, 7-15 査読あり
 6. Ninomiya, H., Hiramatsu, M., Inamura, K., Nomura, K., Okui, M., Miyoshi, T., Okumura, S., Satoh, Y., Nakagawa, K., Nishio, M., Horai, T., Miyata, S., Tsuchiya, E., Fukayama, M., Ishikawa, Y. (2008), "Correlation between morphology and EGFR mutations in lung adenocarcinomas Significance of the micropapillary pattern and the hobnail cell type", *Lung Cancer*, Vol. 63. No. 2, 235-240. 査読あり
 7. Oku, Y., Shimoji, T., Takifuji, K., Hotta, T., Yokoyama, S., Matsuda, K., Higashiguchi, T., Tominaga, T., Nasu, T., Tamura, K., Matsuura, M., Miyata, S., Kato, Y., Yamaue, H., Miki, Y. (2008), "Identification of the molecular mechanisms for dedifferentiation at the invasion front of colorectal cancer by a gene expression analysis", *Clinical Cancer Research*, No. 14, Vol. 22, 7215-7222. 査読あり
 8. Ueda, N., Tada, K., Miyata, S., Koizumi, M., Kuroda, Y., and Iwase, T. (2008), "Identification of Sentinel Lymph Node Location Based on Body Surface Landmarks in Early Breast Cancer Patients", *Epub ahead of print in Breast Cancer* 査読あり
 9. Nunobe, S., Ohyama, S., Miyata, S., Matsuura, M., Hiki, N., Fukunaga, T., Seto, Y., Ushijima, M., and Yamaguchi, T., (2008) "Incidence of gastric cancer in the remnant stomach after proximal gastrectomy.", *Hepatogastroenterology*, No. 55, Vol. 86-87, 1855-1858. 査読あり
 10. Hiki, N., Fukunaga, T., Yamaguchi, T., Ogura, T., Miyata, S., Tokunaga, M., Sano, T. (2009), "Increased Fat Content and Body Shape Have Little Effect on the Accuracy of Lymph Node Retrieval and Blood Loss in Laparoscopic Distal Gastrectomy for Gastric Cancer", *Journal of Gastrointestinal Surgery*, No. 13, Vol. 4, 626-33. 査読あり
 11. Satoh, Y., Hoshi, R., Horai, T., Okumura, Nakagawa, K., Ishikawa, Y., Miyata, S. (2009), "Association of Cytologic Micropapillary Clusters with Lymphatic Spread in Clinical Stage I Lung Adenocarcinomas", *Lung Cancer*, Vol. 64, No. 3, 277-281. 査読あり
- [学会発表](計11件)
1. Isomura, M., Matsuura, M., Nagasaki, K., Ushijima, M., Miyata, S., Noda, T., Miki, Y. (2008), "Genome-wide survey of genetic polymorphisms associated with cytotoxicity of paclitaxel using HapMap cell lines derived from Japanese population", Second JCA-AACR Special Joint Conference. The Latest Advances in Breast Cancer Research
 2. Satoshi Miyata (2008), "Adaptive Model Selection for Nonparametric Density Estimation", Joint Session of Chinese Statistical Association (CSA), Japanese Statistical Society (JSS) and Korean Statistical Society (KSS)
 3. 宮田 敏 (2007), "Adaptive model selection for regression, classification and density estimation", 京都大学経済研究所計量経済学セミナー
 4. 牛嶋大*, 宮田敏, 松浦正明 (2007), 「質量分析計データを用いた判別解析とバイオマーカー探索」2007 年度統計関連学会連合大会
 5. 宮田敏* (2007), 「適応型モデル選択基準に基づく確率密度推定」2007 年度統計関連学会連合大会
 6. Yoshimasa Oku, Yakashi Shimoji, Masaaki Matsuura, Satoshi Miyata, Katsunari Takifuji, Tsukasa Hotta, Yo Katou, Tetsuo Noda, Yoshio Miki (2007), "Gene expression profiles

related with dedifferentiation at the invasion front of colorectal cancer”, 第 66 回日本癌学会総会

7. Yumi Enomoto, Masaru Ushijima, Satoshi Miyata, Masaaki Matsuura (2007), “The System for prediction of phenotype and inference of their regulations mechanism using multivariate analysis”, 第 66 回日本癌学会総会
8. Satoshi Miyata (2007), “Adaptive Model Selection for Regression, Classification and Density Estimation”, 一橋大学科学研究費シンポジウム「統計的モデリングの方法と理論」
9. 宮田 敏 (2007)、「適応型モデル選択基準に基づく回帰分析、判別分析および密度推定」、名古屋統計セミナー
10. 牛嶋 大、宮田 敏、松浦 正明 (2008)、「プロテオームデータ解析の現状および統計的諸問題」2008 年度統計関連学会連合大会
11. 宮田 敏、牛嶋 大、松浦正明 (2008)、「コネクティヴィティマップを用いた潜在的に有効な抗がん剤の遺伝子発現印章に基づく識別」第 67 回日本癌学会総会

6. 研究組織

(1)研究代表者

宮田 敏 (MIYATA SATOSHI)
財団法人 癌研究会 ゲノムセンター
研究員
研究者番号：60360343

(2)研究分担者

なし

(3)連携研究者

なし