

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 5 日現在

機関番号：82626
研究種目：基盤研究(C) (一般)
研究期間：2014～2016
課題番号：26330194
研究課題名(和文) 生物学分野における計測画像の解析手法に関する研究

研究課題名(英文) Multivariate analysis method for biology

研究代表者
渡辺 顕司 (Watanabe, Kenji)

国立研究開発法人産業技術総合研究所・知能システム研究部門・研究員

研究者番号：50571064

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：一般に、生物学分野における画像解析システムは特徴抽出手法と識別手法からなる。ここで、適用する問題設定に合わせた特徴抽出手法を選択・適用できれば、識別性能の向上が期待できる。しかし、数理・情報科学的知見の乏しい生物学系研究者等が、この手法を取捨選択するのは非常に困難である。そこで本研究では、数理・情報科学的知見を持たなくとも識別性能向上効果が期待できる特徴量変換手法を提案した。本研究の提案手法と、既存の特徴量変換手法を医用画像の識別問題に適用・評価したところ、提案手法は安定した識別性能向上効果を示した。

研究成果の概要(英文)：Analysis systems for biological images generally comprise a feature extraction method and a classification method. Task-oriented methods for feature extraction are very effective at improving the classification accuracy. However, it is difficult to utilize such feature extraction methods for versatile task in practice, because few biologists specialize in mathematics and/or informatics to design the task-oriented methods. Thus, in order to improve the usability of these supporting systems, it will be useful to develop a method that can automatically transform the image features of general propose into the effective form toward the task of their interest. In this work, we propose a semi-supervised feature transformation method, which is formulated as a natural coupling of principal component analysis and linear discriminant analysis. Compared with other feature transformation methods, our method showed favorable classification performance in biological image analysis.

研究分野：パターン認識

キーワード：多変量解析 半教師あり機械学習 特徴量変換

1. 研究開始当初の背景

近年、生物学分野では、これまでに広く用いられている細胞染色法の改良や計測技術の各段の進歩により、短時間で多量かつ高精細な計測が可能となり、新規知見の獲得に利用されている。このような計測実験では、培養細胞や実験動物の組織切片中における対象物質（たとえば、タンパク質や核酸等）に関する染色画像を取得し、これらの画像から、任意の状態（たとえば、複数種類タンパク質間の結合状態など）を識別し、病理機序の解析などを行っている。

以上の解析では、研究者が計測データを手動で評価し、学術的研究成果などの先見知識を用いて立てた仮説と比較・検討することで新規知見の獲得を行っている。このような手動データ評価・解析では、研究者への負担が非常に大きなものとなるうえに、作業に多大な時間を要することが多い。さらに、熟練した研究者と学生などの初学者では、評価・解析精度に大きな差が生じてしまうという問題もある。したがって、研究者の負担軽減および評価・解析時間の短縮は、生物学分野における国際競争力向上の観点からも早期の解決を望まれる問題であり、この問題解決に資する研究が、情報工学の分野でも盛んに研究されるようになってきている。

情報工学分野からの生物学的計測画像解析へのアプローチは、画像からの特徴抽出手法と機械学習手法を用いた識別手法[1, 2]や、任意のモデルを用いたヒューリスティックな手法[3]を用いた解析システムの提案・構築である。これらの画像解析手法では、特徴抽出手法に利用者の（統計的・物理的）先見知識を導入するという観点に欠けるもの[2]や、統計的な観点を欠くために汎化性に難があるもの[3]が多い。さらに、これらの手法は、前提として数理的・情報幾何的知見を持っていないければ、好適な特徴抽出・正しい識別結果の解釈が困難な場合が多い[1]。

2. 研究の目的

本研究における申請当初の目的は、手法の利用者が数理・情報幾何的な知見を持たなくても、好適な特徴抽出と、正しい識別（類別）結果の解釈を可能とするため、パターン認識の観点から、解析手法を利用する研究者・技術者が持つ先見知識を考慮したうえで、自動で好適な特徴量を抽出できる、新たな特徴抽出手法を提案する事であった。さらに、利用者が直感的に理解可能な類別結果を出力できる識別手法を提案したのち、これらの特徴抽出・識別手法を統合した計測画像解析手法を検討し、「分子生物学分野における新たな計測画像解析手法の確立」と「実験現場への普及」に資することが、研究の出口として想定されていた目標である。

3. 研究の方法

研究開始当初は、数理的・情報幾何的知見が無ければ適切な利用が困難な特徴抽出手法が主流であった。しかし、多階層型の neural network を用いた機械学習の枠組みによる特徴抽出手法の有用性が Krizhevsky らによって示され[4]、医用画像認識にも応用されるようになった[5]。

以上の機械学習の枠組みによる特徴抽出手法の利点は、多量データで事前に学習したモデルを用いることで、個別の問題ごとに再学習を行わずとも、一定以上の識別性能・汎化性能を持つ特徴量が抽出できる点である。また、上述の条件ならば、利用者が数理的知見を持たずとも、容易に、高い識別性能が達成できる点も当該手法の優れた点である。

一般に、機械学習の枠組みによる特徴抽出手法を識別問題に適用する場合、識別性能のさらなる向上を目指すならば、適用する識別問題に合わせて、多層の neural network の再学習が必須である。しかし、この再学習には、莫大な計算コスト（たとえば、CPU 占有時間や占有メモリ量）と、任意のパラメータ探索が必要であり、再学習を容易に行うのは困難である。また、抽出された特徴量も、問題ごとの先見知識が陽に導入されていないため、意味解釈が困難な場合が多い。

そこで本研究では、申請当初の目的から研究内容を絞り込むことで、上記の有用な特徴量などを、特徴抽出手法の再学習なしで問題ごとに適応させ、識別性能を向上させるための特徴量変換手法を提案した。この提案手法は、基礎的な多変量解析手法へと立ち返り、主成分分析、および線形判別分析をレイリー商の形式で定式化した後、半教師あり機械学習手法の枠組みで統合することで、学習データに関するサンプル分布の分散を最大化し、かつ、（ガン・非ガン等の）クラスラベルを付与された教示ラベル付きデータに関するサンプル分布を判別基準の意味で好適に変換できる手法とした。

本提案手法は、物体認識における認識性能向上効果が見込めることを検証する為、以下の評価実験を実施した。

- 1) 「見え」の異なる画像情報を用いた個人同定問題における同定性能向上効果の検証 [6]
- 2) 医用画像の識別問題における識別性能向上効果の検証 [7]

以上の評価実験では、既存の特徴量変換手法と比較した、提案手法の優位性を確認したほか、提案手法を用いることで、特徴抽出の有

無に関わらず、認識性能の向上が見込めることを確認した。

4. 研究成果

(1) 本研究の成果は、主成分分析を基礎とし、これにクラスラベル付きサンプルに関する判別基準を導入した、半教師あり機械学習手法の枠組みによる特徴量変換手法を新たに提案したことである。我々の提案手法は、グラフ理論から、主成分分析と判別分析を自然な形で統合した定式化に成功しており、そのグラフ構造は入力データのクラスラベルのから決定される。

提案手法の評価関数は、主成分分析、および判別分析に関するラプラシアン行列の重み付き和で表現される。また、この評価関数は、下式に示す通り、全分散行列、および級内分散行列の重み付き和としても表現できる。

$$f(A) = \text{tr} \left(A^T \left[\Sigma - \frac{n_l}{n} \Sigma_w \right] A \right),$$

ここで、 A は特徴量の変換行列であり、 Σ 、および Σ_w は、それぞれ入力データの全分散行列、および入力データのクラスラベル付きサンプルのみで算出した級内分散行列である。 n_l 、および n_u は、それぞれ入力データのクラスラベル付きサンプル数、およびクラスラベルなしサンプル数であり、入力データの全サンプル数は $n = n_l + n_u$ となる。

以上の評価関数から、本提案手法は、以下の最適化問題を解くことで、変換行列 A を求める。

$$\begin{aligned} A^* &= \arg \max_A f(A), \\ \text{s.t. } &\alpha \frac{n_u}{n} A^T A + \frac{n_l}{n} A^T \Sigma_w A = I. \end{aligned}$$

ここで、 α は任意の係数となる。本提案手法は、 $\alpha = 1$ の場合、「全サンプルにクラスラベルが付与される場合は、判別分析と同一」となり、また、「全サンプルにクラスラベルが付与されない場合は、主成分分析と同一」となる。

制約条件における左辺第一項と第二項を見ると、 $\alpha = 1$ とした場合、各項の重みに関して不均一さを生じてしまい、実環境で計測したデータに対する、好適な変換行列推定が困難になる可能性がある。そこで本研究では、クラスラベル付きサンプル数に起因する制約条件中の各項の不均一さを緩和するため、 $\alpha = \text{tr}(\Sigma) / m$ とした[7]。ここで m は特徴次元数である。

(2) 生物学分野における画像解析を実施する場合、当該解析を実施する機関ごとに、計測装置の機種（計測機種）が異なるのは、十分に予想できる事である。たとえば、図1に示す通り、同一の対象を異なるメーカー製の計測装置を用いることで、その色合いや精細さ等が異なる場合が多々発生する。

本研究では、「提案手法を適用することで、使用する計測機種によらず、安定した画像解析を実現できる事」を検証するため、図1のデータを用いて、計測機種ごとに、細胞の正常/異常を識別する識別問題に適用した。上記の識別問題では、AlexNet [4]を用いて入力画像から特徴抽出（表1 CNN）を行い、抽出した特徴量の変換に、提案手法（KSCA）を適用する（表1 CNN + KSCA）ことで、識別性能の向上効果、および識別率の安定化に関する検証を行った。

表1から、提案手法を適用することで、計測機種によらず、5%程度の識別率向上が可能であり、かつ計測機種間の識別率変化を0.1%程度にまで減少する事に成功した。これは、既存の画像解析手法になどに、提案手法を適用することで、さらなる解析精度向上が見込めることを示している。

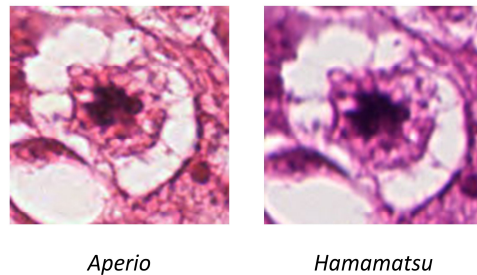


図1 異なる機種で計測した組織画像例
MITOS-ATYPIA-14 dataset [8]

表1 計測機種ごとの識別率[7]

	Aperio	Hamamatsu
CNN	0.718 (± 0.012)	0.723 (± 0.008)
CNN + KSCA	0.776 (± 0.015)	0.775 (± 0.012)

<引用文献>

- [1] A. Yaguchi, et al., *Proceedings of ICIP2011*, pp. 1641-1644, (2011).
- [2] H. Nosato, et al., *Proceedings of MIRU*, (2010).
- [3] C. Demir, et al., *TECHNICAL REPORT, RENSSELAER POLYTECHNIC INSTITUTE, DEPARTMENT OF COMPUTER SCIENCE*, TR-05-09, (2005).
- [4] A. Krizhevsky, et al., *Proc. of NIPS*, (2012).
- [5] CK. Shie, et al., *Proc. of Int. conf. of EMBC*, pp.711-714, (2015).
- [6] K. Watanabe and T. Wada, *Proc. of SMC2015*, pp.3011-3016, (2015).
- [7] K. Watanabe, et al., *PLOS ONE*, 11(12), (2016), e0166413
- [8] MITOS-ATYPIA-14 dataset. Accessed: <https://grand-challenge.org/site/mitos-atypia-14/>

(2)研究分担者

和田 俊和 (WADA, Toshikazu)
和歌山大学・システム工学部・教授
研究者番号 : 00231035

5 . 主な発表論文等

〔雑誌論文〕(計 2件)

K. Watanabe, T. Kobayashi, and T. Wada,
“Semi-Supervised Feature Transformation
for Tissue Image Classification,” *PLOS
ONE*, 11(12), 2016, e0166413, doi:
10.1371/journal.pone.0166413, peer review
K. Watanabe and T. Wada,
“Semi-supervised Component analysis,”
*Proc. of 2015 IEEE Int. Conf. on Systems,
Man, and Cybernetics (SMC2015)*, 2015,
pp.3011-3016, doi: 10.1109/SMC.2015.524,
peer review

〔学会発表〕(計 1件)

K. Watanabe, and T. Wada,
“Semi-supervised Component analysis,”
*2015 IEEE Int. Conf. on Systems, Man, and
Cybernetics*, 11 Oct. 2015, Hong Kong
(China).

出願状況 (計 0件)

取得状況 (計 0件)

〔その他〕

<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0166413>

6 . 研究組織

(1)研究代表者

渡辺 顕司 (WATANABE, Kenji)
産業技術総合研究所・知能システム研究部
門・研究員
研究者番号 : 50571064