

《技術報告》

大容量の画像データを扱う主成分分析の 高速で正確な計算方法

袁島 聡* Robert A. KOEPPÉ*

要旨 主成分分析は、核医学における様々な画像検査に応用されてきた。この技術報告では、より大容量の画像データを対象とした主成分分析における、固有値および固有ベクトルの高速で正確な計算方法を記述する。この方法では、分散共分散あるいは相関行列を計算するために、もととなるデータのデータ変換および転置行列の計算を行う。実際の画像データと普及型ワークステーションを用いてこの方法を試行したところ、通常の方法に比べ、より高速でかつ十分な精度を示すことを確認した。

(核医学 33: 431-434, 1996)

I. はじめに

主成分分析は、観測されたデータをより少ない指標で表現し、それら指標をもってデータを解釈しようとする方法である。主成分分析および近縁の因子分析は、比較的早くより核医学画像の解析に用いられ、当初はダイナミック収集された腎、心、胆嚢スキャンなどのブレイナー画像において、投与放射能の臓器内における経時的変化、トレーサ動態解析を目的として使用された¹⁻⁵⁾。さらに PET や SPECT など核医学断層像検査法が出現してからは、例えば脳局所の機能的連絡や変動パターン等を解析するために、関心領域により抽出したデータに対して、主成分分析や因子分析が

用いられた⁶⁻¹⁰⁾。しかし、最近のワークステーションの普及に伴い、大きな3次元画像データを対象として画素ごとに、これら解析法を試みる報告がなされている¹¹⁻¹³⁾。主成分分析では、対象とするデータを含む行列の固有値および固有ベクトルを計算する必要がある。しかし大量の画素値を検討対象とした場合は、使用する計算機の記憶量や計算速度に限りがあるため、通常の方法で解を求めることはきわめて非合理的かあるいは不可能な場合が多い。そのため代わりとなる解法がいくつか報告されている^{12,13)}。本技術報告では、転置行列を用いた方法を明確にし、その実際の計算における問題点を述べ、実際に画像データを用いてその計算速度および精度を検討した。

II. 対象と方法

ここに画像データを含む $n \times p$ の行列 **I** を考える。 n は全画像数、 p は1画像中の全画素数を示し、行列の各成分は各画像中の各画素における測定値を示すものとする。画像は2次元あるいは3次元のいずれでも良いが、1画像中の全画素は行列 **I** の1つの行に一定の順で再配列されている。実際の例としては、 128×128 画素よりなるブレイナー画像 ($p=16384$) を放射性薬剤投与後経時

* Division of Nuclear Medicine, Department of Internal Medicine, The University of Michigan Medical Center, USA

受付: 7年11月6日

最終稿受付: 8年1月10日

別刷請求先: Satoshi Minoshima, M.D., Ph.D.

Division of Nuclear Medicine

Department of Internal Medicine

The University of Michigan Medical Center

B1G412 University Hospital

Ann Arbor, Michigan 48109-0028

U.S.A.

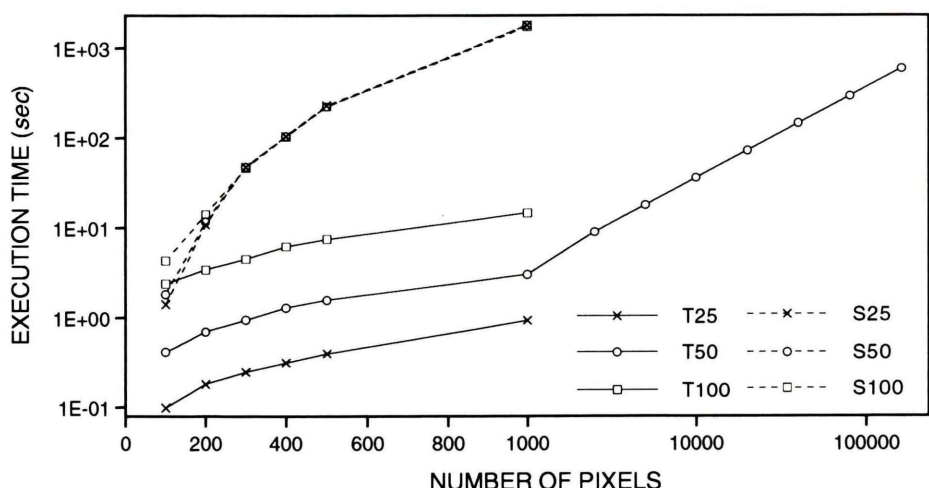


Fig. 1 Execution time of two different algorithms to calculate principal components: Standard approach (S, dash line) and the use of transposed matrix (T, solid line). The numbers following T and S denote numbers of data sets (n) included in the calculation. The horizontal axis (logarithmic scale for $p > 1000$) represents the number of pixels (p). The vertical axis represents execution time in seconds (logarithmic scale). It is obvious that T is faster than S when $p > 200$.

的に 20 回 ($n=20$) 撮像したデータ, 64×64 画素 $\times 20$ スライス ($p=81920$) よりなる SPECT 画像を経時的に 12 回 ($n=12$) 撮像したデータ, あるいは解剖学的に標準化された 128×128 画素 $\times 60$ スライスの PET 糖代謝画像中に含まれる灰白質の画素 (e.g., $p=100000$) の被験者 30 人 ($n=30$) のデータなどである。この行列 **I** は上記に示すごとく、きわめて大きい p に対して比較的小さい n を持つのが特長である。行列 **I** の分散共分散行列 **Σ** あるいは相関行列 **C** は $p \times p$ の行列となるが、主成分分析とは、すなわちこの行列 **Σ** あるいは **C** の固有値 λ および固有ベクトル **E** を求めることである。しかしながら、きわめて大きい $p \times p$ 行列 **C** の固有値および固有ベクトルを通常の方法で求めることは、その膨大な計算量から現実的には不可能である場合が多い。そこで行列 **I** の転置行列 **I'** を考える。**I'** は $p \times n$ 行列であり、その分散共分散行列 **Σ** あるいは相関行列 **c** は $n \times n$ 行列となる。この行列 **Σ** あるいは **c** は固有値 λ を持ち、固有ベクトルを **e** とした場合、 $\mathbf{E} = \mathbf{I} \mathbf{e} / \sqrt{\lambda}$ の関係が成り立つ¹³⁾。すなわち、行列 **Σ** あるいは **C** の固有値

λ および固有ベクトル **E** は、行列 **I** および分散共分散行列 **Σ** あるいは相関行列 **c** の固有値 λ および固有ベクトル **e** から求められることがわかる。

実際にこの転置行列を用いて固有値 λ および固有ベクトル **E** を求める場合、分散共分散行列あるいは相関行列を求める前にあらかじめ行列 **I** の各成分を変換しておく必要がある。分散共分散行列を用いた主成分分析を行う場合には、行列 **I** の各列 j の平均値 μ_j ($\mu_j = \sum a_{ij}/n, i=1, \dots, n$) をその列の各成分 a_{ij} から減算した上で行数 $n-1$ の平方根で除する $((a_{ij} - \mu_j)/\sqrt{n-1})$ 。この結果を仮に行列 **J** とする。相関行列を用いる場合には、さらにその行列 **J** の各列 j に含まれる成分 b_{ij} の二乗和 d_j を求め ($d_j = \sum (b_{ij} \times b_{ij}), i=1, \dots, n$)、その平方根でその列 j の各成分を除する $(b_{ij}/\sqrt{d_j})$ 。この結果を行列 **K** とする。これら行列 **J, K** より、上記の方法で主成分分析を行う。例えば相関行列を用いる場合には、行列 **K** より計算した相関行列 **C** は $\mathbf{K}'\mathbf{K}$ で $p \times p$ の行列となるが、**K** の転置行列 **K'** を用いて相関行列 **c** を計算すると、行列 **c** は $(\mathbf{K}')'\mathbf{K}'$ より $\mathbf{K}\mathbf{K}'$ となる ($n \times n$ 行列)。この行列 **c** より固

有値 λ および固有ベクトル \mathbf{e} を求め、 $\mathbf{E} = \mathbf{K}\mathbf{e} / \sqrt{\lambda}$ により相関行列 \mathbf{C} の固有ベクトル \mathbf{E} を計算できる。同様に分散共分散行列を用いる場合には、行列 \mathbf{J} より計算した分散共分散行列 $\mathbf{\Sigma} = \mathbf{J}'\mathbf{J}$ を求めて分析する代わりに、 \mathbf{J} の転置行列 \mathbf{J}' を用いて $n \times n$ の分散共分散行列 $\mathbf{\Sigma} = \mathbf{J}\mathbf{J}'$ を求めて解析し、 $\mathbf{E} = \mathbf{J}\mathbf{e} / \sqrt{\lambda}$ により $\mathbf{\Sigma}$ の固有ベクトル \mathbf{E} を求める。

上記の転置行列を用いた主成分分析を実際の画像データを用いて試験し、通常の方法を用いた場合と比較した。内部メモリ 64 M バイトを有する SPARCstation 2 (Sun Microsystems, Mountain View, CA) を使用し、プログラムには UNIX 標準 C 言語、計算には単精度浮動小数を用いた。主成分分析は相関行列を使用した。固有値および固有ベクトルは、いずれの方法においても Householder 法による行列変換に引き続く QL アルゴリズムを用いて求めた¹⁴⁾。画像データとしては、100 症例の解剖学的標準化された 128×128 画素 \times 60 スライスの脳糖代謝 PET 画像を用いた¹⁵⁾。画像上で頭部全体を含むおよびその輪郭を決定し、その内部に含まれる画素を頭頂部より必要画素数 p だけ用いた。画像 (症例) 数 n は 25, 50 および 100、画素数 p は 100 から 160000 を想定し ($n < p$)、転置行列を用いた方法と通常の方法による計算速度および精度を比較した。

III. 結 果

いずれの n および p の組み合わせにおいても、転置行列を用いた方法は、通常の方法に比較して高速に固有値および固有ベクトルを計算し得た (Fig. 1)。(n, p) = (100, 1000) の場合、通常の方法では約 29 分の時間を要するのに対し、転置行列を用いた方法では約 15 秒でありその比は約 119 である。 p がさらに大きくなった場合に、さらにその差が広がるのは明らかであるので、 $p > 1000$ に関しては $n = 50$ で転置行列を用いた方法のみ検討した。その結果 $p = 160000$ で約 10 分を要するのみであった。これに対して求めた固有値から計算した全累積寄与率の 100% からのずれを誤差の

指標とした場合、 $p < 1000$ では通常の方法および転置行列を用いた方法のいずれも $5 \times 10^{-5}\%$ 以下であった。 p が大きくなるにつれ誤差は増加したが、転置行列を用いた方法で (n, p) = (50, 2000) では約 $7 \times 10^{-5}\%$ 、(50, 40000) で約 $2 \times 10^{-3}\%$ 、(50, 160000) で約 $4 \times 10^{-2}\%$ 程度であった。 p がこのように大きくなった場合、使用したワークステーションの仕様では通常の方法を実行することは不可能であった。

IV. 考察および結語

以上に示されたように、大きな画像データを対象とした固有値および固有ベクトルの計算において、転置行列を用いた方法は計算速度および精度共に満足のいくものであった。この方法により、画像そのものを対象とした主成分分析の応用が、通常のワークステーションを用いてきわめて容易になる。本報告ではデータ変換および転置行列を用いた方法が、通常の方法に比べてきわめて高速に解を求めることを示したが、転置行列を用いた方法は帰納的アルゴリズムを用いた方法¹²⁾に比較しても高速であることが示唆されている¹³⁾。

大きな画像データの主成分分析の計算に転置行列を用いる方法は以前より示唆されており¹¹⁾、最近の報告ではより明確な論理の証明がなされている¹³⁾。しかし実際にこの方法を用いるにあたって、もとの行列 \mathbf{I} から単に転置行列 \mathbf{I}' を求めその分散共分散行列あるいは相関行列から解を求めた場合¹³⁾、その解は行列 \mathbf{I} より通常の方法で求めたものと一致しない。本報告で述べたように、分散共分散行列あるいは相関行列を計算する前に行列 \mathbf{I} を変換する必要がある。MATLAB などの計算言語を用いてプログラムを導入する場合は、注意する必要があると思われる。

本報告では、画像を対象とした主成分分析における固有値および固有ベクトルを、高速かつ正確に計算する方法を示し検証した。核医学画像を対象とした主成分分析の応用により、より新しいデータ解釈法の発展が望まれる。

文 献

- 1) Barber DC: The use of principal components in the quantitative analysis of gamma camera dynamic studies. *Phys Med Biol* **25**: 283–292, 1980
- 2) Oppenheim BE, Appledorn CR: Functional renal imaging through factor analysis. *J Nucl Med* **22**: 417–423, 1981
- 3) 中嶋憲一, 山田正人, 松平正道, 分校久志, 利波紀久, 久田欣一, 他: 心ファントムおよび数学的モデルによる因子分析法の検討. *核医学* **23**: 1423–1434, 1986
- 4) Samal M, Surova H, Karny M, Marikova E, Michalova K, Dienstbier Z: Enhancement of physiological factors in factor analysis of dynamic studies. *Eur J Nucl Med* **12**: 280–283, 1986
- 5) Oster ZH, Som P, Bazin JP, Di Paola M, Raynaud C, Atkins HL, et al: The role of factor analysis in the evaluation of new radiopharmaceuticals. *Int J Rad Appl Instrum [B]* **16**: 85–89, 1989
- 6) Riege WH, Metter EJ, Kuhl DE, Phelps ME: Brain glucose metabolism and memory functions: age decrease in factor scores. *J Gerontol* **40**: 459–467, 1985
- 7) Volkow ND, Brodie JD, Wolf AP, Gomez-Mont F, Cancro R, Van Gelder P, et al: Brain organization in schizophrenia. *J Cereb Blood Flow Metab* **6**: 441–446, 1986
- 8) Szabo Z, Camargo EE, Sostre S, Shafique I, Sadzot B, Links JM, et al: Factor analysis of regional cerebral glucose metabolism rates in healthy man. *Eur J Nucl Med* **19**: 469–475, 1992
- 9) McLaughlin T, Steinberg B, Christensen B, Law I, Parving A, Friberg L: Potential language and attentional networks revealed through factor analysis of rCBF data measured with SPECT. *J Cereb Blood Flow Metab* **12**: 535–545, 1992
- 10) Frouin F, Cinotti L, Benali H, Buvat I, Bazin JP, Millet P, et al: Extraction of functional volumes from medical dynamic volumetric data sets. *Comput Med Imaging Graph* **17**: 397–404, 1993
- 11) Houston AS, Kemp PM, Macleod MA: A method for assessing the significance of abnormalities in HMPO brain SPECT images. *J Nucl Med* **35**: 239–244, 1994
- 12) Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ: Functional connectivity: the principal-component analysis of large (PET) data sets. *J Cereb Blood Flow Metab* **13**: 5–14, 1993
- 13) Weaver JB: Efficient calculation of the principal components of imaging data. *J Cereb Blood Flow Metab* **15**: 892–894, 1995
- 14) Press WH, Teukolsky SA, Vetterling WT, Flannery BP: *Numerical Recipes in C*. Cambridge University Press, Cambridge, 1988, pp. 368–381
- 15) Minoshima S, Koeppe RA, Frey KA, Kuhl DE: Anatomic standardization: linear scaling and nonlinear warping of functional brain images. *J Nucl Med* **35**: 1528–1537, 1994

Summary

Fast and Accurate Numerical Method for Principal Components Analysis Dealing with Large Image Data Sets

Satoshi MINOSHIMA and Robert A. KOEPE

Division of Nuclear Medicine, Department of Internal Medicine, The University of Michigan

The principal components analysis has been applied to various imaging studies in nuclear medicine. This technical report describes a fast and accurate numerical method of calculating eigenvalues and eigenvectors in the principal components analysis dealing with larger image data sets. The method employs both data transformation and matrix transpose of original data

sets to calculate a variance-covariance or correlation matrix. The method was tested on actual image data sets using a common workstation, confirming faster execution time and efficient accuracy in comparison to a standard method.

Key words: Principal components analysis, Computer-assisted image analysis.