様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２６ 年　 ６ 月　 ５ 日現在

機関番号：１２６０８
研究種目：基盤研究(B)
研究期間：2011～2013
課題番号：２３３０００６９
研究課題名（和文）拡張カーネル法による信号多様体の時空間計量表現とその応用

研究課題名（英文）signal manifold by extended kernel method and its application

研究代表者
山下　幸彦（Yukihiko, Yamashita）

東京工業大学・理工学研究科・准教授

研究者番号：９０２２０３５０

交付決定額（研究期間全体）：（直接経費）　15,900,000円、（間接経費）　4,770,000円

研究成果の概要（和文）：　信号処理やパターン認識の性能向上のためには，信号の分布の偏りを反映できる距離構造を求める方法と，それに基づいた信号処理・パターン認識の枠組みが必要である。
その理念のもと，Mahalanobis 計量理論の深化と，カーネル法による計算法を開発し，事後確率を求める必要がない，最大時後確率法と同等な識別法を開発した。また，確率密度関数間のL2距離を直接推定する手法を考案し，Stiefel多様体上でその標本平均を効率的に求めるアルゴリズムを開発すると共に，基底数を増やさずに効率的な特徴抽出が可能な適応部分カーネル主成分分析を開発した。そして，開発した手法の有効性を計算機実験によって確認した。

研究成果の概要（英文）：　In order to improve the performance of signal processing and pattern recognition, it is necessary to obtain a proper metric for signals and patterns and develop processing and recognition methods based on the metric.
According to the concept, we improve the theory of Mahalanobis metric, develop a calculation method of Mahalanobis metric using kernel functions, and provide criteria and calculation methods that can obtain a equivalent result with the maximum posteriori method without estimating a posteriori probability. We also propose a method to directly estimate the L2 distance between two probability density functions and show its properties with respect to convergence and bias. We develop an algorithm to calculate the sample average on Stiefel manifold, and the adaptive kernel principle component analysis that can extract features with a small number of basis functions. The advantages of developed methods are shown by experimental results.

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：時空間計量　Mahalanobis計量　最小二乗確率的分類器　多様体上のアンサンブル学習　マルチカーネル適応フィルタ　制約条件付き最大事後確率識別器

1．研究開始当初の背景

信号処理やパターン認識で所望の出力や結果を得るためには，信号サンプル間の「距離」を適切に定めることが重要である。そして，一般に信号の分布には何らかの「偏り」があるため，古典的な Euclid 距離の枠組みを越え，その偏りを反映できる距離構造を求める方法と，それを考慮した信号処理・パターン認識の枠組みが必要とされていた。

筆者らが提案した (幾何学的) 局所等方独立[1]に基づく Mahalanobis 計量は，信号の確率分布をもとにした適切な計量を求めることができる。$p(x,y)$ を 2 次元 Euclid 空間の確率密度関数とすれば，(幾何学的) 局所独立を，

$$\frac{\partial}{\partial y}\frac{1}{p}\frac{\partial p}{\partial x}=0 \tag{1}$$

で定義した。これを一般座標系における等方的な (幾何学的) 局所独立に拡張し，座標変換に共変な方程式に変形する。$g_{\mu\nu}$ を計量テンソル，g をその行列式，∇_μ を共変微分，ある l をスカラ関数とするとき，その方程式は，

$$\nabla_\mu\nabla_\nu\frac{p}{\sqrt{g}}=lg_{\mu\nu} \tag{2}$$

で与えられる。これは元々は多様体の計量が与えられた時に確率密度関数を与えるものであった。一方この式は，確率密度関数が与えられた時に計量テンソルを求めることによって，Mahalanobis 計量を求めることができる。未定のスカラー l は，Mahalanobis 計量を定義するための元にする空間が，曲率が 0 でフラットな Euclid 空間の場合は $l=-1$，正の定曲率を持つ超球面の場合は $l=-\log p/\sqrt{g}$，負の定曲率を持つ Lobachevsky 空間の場合は $l=\log p/\sqrt{g}$ とすれば良い。しかしながら，局所独立に関して，$p(x)$, $p(y)$ を x, y の周辺確率密度関数とすれば, 2 次元 Euclid 空間では，(幾何学的) 局所独立は確率論における独立 $p(x,y)=p(x)p(y)$ と等価になるが，一般の空間ではその性質が明らかではなかった。時空間信号を扱う場合に向けての，Mahalanobis 計量理論の深化がまだ不十分であった。また，Mahalanobis 計量を実際に計算するためには，信号空間を格子点に分割して近似しなければならなかった。従って，高次元空間では莫大な計算量が必要となっていた。

パターン識別に関して，最大事後確率法は学習パターン数が無限大の場合，識別率が最良になるが，事後確率を求める評価基準は，Strictly Sense Bayesian なものに限られ，例えば，線形計画法で解くことができる評価関数は存在しないなど，評価基準の自由度が低かった。また，事後確率を求める精度が低いという問題があった。さらに，Fisher 線形判別は，パターンが正規分布に従う場合においても，識別のために最適な特徴を与えない。その問題を解決するために，Chernoff 距離に基づく評価基準が与えられているが，それは近似的的なものであるという問題があった。

そして，カテゴリ推定の信頼度や確率密度間の距離を直接的に求めることが可能で，高速に計算することができる識別法が存在していなかった。

構造化行列は，行列の最適化問題や医療画像，適応 フィルタリング，計算器科学などの分野で広く使われている。構造化行列のデータの集合を統計的に扱う場合に，経験的平均はその集合を特徴付ける有用な方法のひとつである。このとき，データが持つ構造を保持する必要がある。例えば，データとなる行列が必ず直交行列になるという条件がある場合に，単純に算術平均を求めると，直交行列ではなくなってしまうという問題がある。直交行列のような Lie 群の場合は計算法が提案されているが，Stiefel 多様体のような非 Lie 群の場合は，経験的平均の効率的な演算法が確立がされていなかった。

脳信号などの多次元の時系列データから特徴抽出を行う場合に，従来の手法では，あらかじめすべての標本を用意しておく必要があったり，時間の経過とともに基底の数が増大するという問題があった。

2．研究の目的

(1) 確率論において，$p(x,y|z)$ を z のもとでの x,y の同時条件付き確率密度関数，$p(x|z)$，$p(y|z)$ をそれぞれ z のもとでの x，y の条件付き確率密度関数とすれば，x と y が z に関して局所独立であるとは，

$$p(x,y|z)=p(x|z)p(y|z)$$

によって定義される。または，確率測度論では，確率測度 P のもとで事象 A と事象 B が事象 C に関して局所独立であるとは，

$$\frac{P(A\cap B\cap C)}{P(C)}=\frac{P(A\cap C)}{P(C)}\frac{P(B\cap C)}{P(C)} \tag{3}$$

と定義される。Mahalanobis 計量の理論の深化のために，(幾何学的) 局所等方独立と確率論の局所独立との関係性を明かにする。また，他の物理的な性質との関連も求め，時空間信号の Mahalanobis 計量を

[1]本研究が始まる前は，幾何学的局所独立が確率論における局所独立とは無関係に定義されていため,「幾何学的」という単語を付加していたが，本研究の成果によって，幾何学的局所独立が局所独立の概念に含まれることがわかったため，現在では「幾何学的」を外して使っている。しかしながら，本研究の以前のことを表す場合に，括弧付きで「幾何学的」を付加して用いている。

定義付けするための礎とする。さらに，Mahalanobis 計量の計算において，計量テンソルをカーネル法を使って表現することによって，高次元空間においても Mahalanobis 計量の計算ができるようにする。

(2) 事後確率を求めることなく，最大事後確率法と等価な認識率を与えることができ，効率的な計算が可能な識別法の理論を確立する。すなわち，x をパターン，c をカテゴリとし，$W(x,c)$ を識別関数とするとき，

$$\operatorname*{argmax}_c p(c|x) = \operatorname*{argmax}_c W(x,c) \tag{4}$$

となる $W(x,c)$ を与えることができる評価基準を求める。さらに，外れ値に頑健な識別法や，パターン分布の局所的な変化に対応できる識別法に拡張する。また，Fisher 線形判別による特徴抽出を改良をするために，評価基準に近似を使わない Chernoff 距離に基づく特徴量を与える方法を開発する。

(3) 最小二乗確率的分類法に基づいて，カテゴリ推定の信頼度を直接的に求めることが可能で，確率密度関数間の距離を推定できる識別法を開発する。また，マルチラベル分類に対応できるようにする。

(4) データとなる構造化行列が，Stiefel 多様体に含まれる場合に，その多様体上で経験的平均を効率的に求めることができる方法を開発する。

(5) 基底を効率的に選択することによって，基底数の増加を抑制ができる，適応カーネル主成分分析法を開発し，脳信号処理の性能を向上させる。

３．研究の方法

(1) Mahalanobis 計量の理論の深化のために，物理的な性質である局所平衡と (幾何学的) 局所独立の等価性を示す。その上で，確率論における局所独立と局所平衡の等価性を示すことによって，確率論における局所独立と (幾何学的) 局所独立の等価性を証明し，時空間信号の Mahalanobis 計量を定義付けするための礎とする。さらに，Mahalanobis 計量の計算において，計量テンソルをカーネル法を使って表現することによって，高次元においても Mahalanobis 計量の計算ができるようにする。

(2) 識別関数の評価基準を，最大事後確率法の評価基準から，標本点数が無限大の場合に最大事後確率法と等価な結果を与えるものに範囲を広げる。このことによって，評価基準が Strictly Sense Bayesian なものに限る必要がなくなり，より広い範囲から識別のためにより適する評価基準を考案する。最大事後確率を直接求める必要がなくなるため，識別率の向上が期待できる。また，効率的な計算手段が存在する，線形計画法などによって計算することができる識別法を開発する。さらに，その評価基準に制約条件が正則化の役割を果たすこと，その項の重み関数と既存の識別法との関係を明かにするとともに，新たに考案した重み関数により外れ値に頑健な識別法, 拡張カーネル法を使った局所的な変化に対応できる識別法を開発し，それらの性能を実験的に評価する。そして，Fisher 線形判別を求める一般化固有値問題に付加項を追加し，繰り返し計算によって得られる Chernoff 距離に基づく特徴量抽出法を開発し，計算機実験によって得られた特徴量を評価する。

(3) 最小二乗確率的分類法は確率的な計算に基づいて分類を行ことが可能で，非常に高速に学習を行うことができるという特徴を持っている。まず，様々な機械学習手法の理論的基盤を整備するために，いくつかの種類の最小二乗確率的分類法に対して，統一的な形で再定式化を行い，その新しい定式化のもと，新しい識別法を開発する。そして，それらの性能を実験的に評価する。

(4) Lie 群型多様体上における平均演算法を，コンパクト Stiefel 多様体における平均演算に拡張する。具体的には，Stiefel 多様体上の点を接空間へ lifting と呼ばれる写像で写し，接空間上で計算した算術平均を retraction と呼ばれる逆写像で，再び Stiefel 多様体へ戻す。さらに，この手法の計算量を削減する方法を開発する。そして，それらの性能を実験的に評価する。

(5) 適応カーネル主成分分析法を開発する。また，適応的かつ計算量を抑えた基底を選択するアルゴリズムを導入し，基底の数を抑えながら特徴抽出を効率的に行なう。そして，それらの性能を実験的に評価する。

４．研究成果

(1) 2 次元 Euclid 空間における確率密度関数 $p(x,y)$ の局所平衡を，座標の微小変化 dx と dy に対して，

$$p(x+dx, y+dy)p(x,y)$$

$$= \quad p(x+dx,y)p(x,y+dy) \tag{5}$$

で定義する。これは例えば，速度 (x,y) と速度 $(x+dx,y+dy)$ の2つの粒子が衝突して，速度 $(x+dx,y)$ と速度 $(x,y+dy)$ の粒子なる確率と，その逆過程の確率が等しく，平衡していて分布が時間的に変わらないことを意味している。式 (5) を座標の微小量の二次まで展開して整理すれば，

$$p\frac{\partial^2 p}{\partial x \partial y} = \frac{\partial p}{\partial x}\frac{\partial p}{\partial y}$$

が得られ，これは式 (1) に変形できる。また，等方性の仮定を加えた局所等方平衡を考え，その方程式を一般座標変換に対して共変な形に変形することによって，(幾何学的) 局所等方独立と局所等方平衡が同値な概念であることを証明した。

次に，確率測度による局所独立と局所平衡を結びつける。式 (3) は，

$$P(A \cap B \cap C)P(C) = P(A \cap C)P(B \cap C) \tag{6}$$

と書くことができる。dx_1, dx_2, dy_1, dy_2 $(dx_1 \gg dx_2 > 0, dy_1 \gg dy_2 > 0)$ を十分小さい正の実数として，事象 A, B, C を，

$$\begin{aligned} A &= [x+dx_1, x+dx_1+dx_2] \\ &\quad \times [y, y+dy_1+dy_2] \\ B &= [x, x+dx_1+dx_2] \\ &\quad \times [y+dy_1, y+dy_1+dy_2] \\ C &= [x, x+dx_1+dx_2] \\ &\quad \times [y, y+dy_1+dy_2] \end{aligned}$$

で定義される長方形の領域とする。このとき，

$$\begin{aligned} A \cap B \cap C &= [x+dx_1, x+dx_1+dx_2] \\ &\quad \times [y+dy_1, y+dy_1+dy_2] \end{aligned}$$

となる。確率密度関数を用いて，式 (6) をそれぞれの確率の dx_1, dx_2, dy_1, dy_2 の二次近似で表せば，

$$\begin{aligned} & p(x+dx_1, y+dy_1)dx_2dy_2 \\ & \cdot p(x,y)(dx_1+dx_2)(dy_1+dy_2) \\ = \; & p(x+dx_1, y)dx_2(dy_1+dy_2) \\ & \cdot p(x, y+dy_1)(dx_1+dx_2)dy_2 \end{aligned}$$

となる。この式を，$dx_2dy_2(dx_1+dx_2)(dy_1+dy_2)$ で割れば，

$$\begin{aligned} & p(x+dx_1, y+dy_1)p(x,y) \\ = \; & p(x+dx_1, y)p(x, y+dy_2) \end{aligned}$$

となり，式 (5) が得られる。このことにより，幾何学的局所独立と確率論における局所独立を結びつけることができた。この関係を使い，局所的に表現された変化 (作用素) に対する独立に関して研究を深化させている。さらに，カーネル関数を使って計量テンソルを表現して，Mahalanobis 計量を求める手法を開発した。

(2) 事後確率を推定することなく，標本点数が無限大の場合に最大事後確率法と等価な結果を与える制約条件付き最大事後確率法を考案した。これは制約条件と凸関数を使った評価式で構成され，制約条件として一次と二次のものを具体的に与えた。E_x で期待値を表すことにする。基本的な評価式は，一次の場合，制約条件

$$\sum_c E_x W(x,c) \quad = \quad 1 \tag{7}$$

$$W(x,c) \quad \geq \quad 0 \tag{8}$$

のもとで，

$$\sum_y E_x p(c|x) \min(W(x,c), 1) \tag{9}$$

を最大化するものである。二次の場合は，式 (7) を，

$$\sum_c E_x |W(x,c)|^2 = 1$$

とするだけである。$W(x,c)$ のモデルをカーネル関数によって与えることによって，この評価基準の具体的な解を求めることができる。

この評価基準の特徴は，二次の場合，サポートベクターマシンと類似したものになっているが，サポートベクターマシンではハイパーパラメータが，カーネル関数のためと正則化のために2つ必要であるのに対し，提案手法ではハイパーパラメータが本質的には，カーネル関数のためのパラメータだけであることである。そのため，提案した評価基準にはパラメータ選択が容易であるという利点が存在する。

カーネル関数で表した $W(x,c)$ を求める方法を与えた。この場合，評価基準全体は凸最適化問題となるため，極大点が存在せず，容易に解くことができることを示した。そして，一次の場合は線形計画法で解くことができること示し，二次の場合は主双対内点法によって解を与える方法を示した。また，上の評価基準の基本形において，最大化する関数を識別関数の差にする方法，制約条件に重み関数を与える方法を示した。そして，二次の場合はその制約条件の重み関数の選び方によって，それがサポートベ

クターマシンや最小二乗学習と等価になること，外れ値に頑健な制約条件の重み関数の与え方を示した。

さらに，マルチカーネルを適用した制約条件付き最大事後確率法，拡張カーネル法を適用したサポートベクターマシンを開発し，パラメータ数の削減やパターン分布の局所的性質に適応することを可能にした。そして，ベンチマークデータや脳信号を使って，提案手法の有効性を示した。

Fisher 線形判別では，クラス間分散を $\boldsymbol{\Sigma}_{\mathrm{B}}$，クラス内分散を $\boldsymbol{\Sigma}_{\mathrm{W}}$ とすれば，その特徴量 $\boldsymbol{v}$ は，一般化固有値問題

$$\boldsymbol{\Sigma}_{\mathrm{B}}\boldsymbol{v} = \lambda\boldsymbol{\Sigma}_{\mathrm{W}}\boldsymbol{v} \tag{10}$$

を解くことによって与えることができる。$\boldsymbol{v}_n$ を，n 回目の繰り返し計算によって得られる，Chernoff 距離に基づいた特徴量とする。このとき，$\boldsymbol{v}_{n+1}$ が一般化固有値問題

$$\left(\boldsymbol{\Sigma}_{\mathrm{B}} - c_1^{(n)}\boldsymbol{\Sigma}_1 - c_2^{(n)}\boldsymbol{\Sigma}_2\right)\boldsymbol{v}_{n+1} = \lambda\boldsymbol{\Sigma}_{\mathrm{W}}\boldsymbol{v}_{n+1} \tag{11}$$

を解くことによって得られることを示した。ここで，実数 $c_1^{(n)}, c_2^{(n)}$ は，$\boldsymbol{v}_n$，クラス間分散，クラス内分散，それぞれのカテゴリー分散によって計算される。また，計算機実験によって，Fisher の線形判別や既存の手法よりも，良い特徴量を求めることができることを示した。

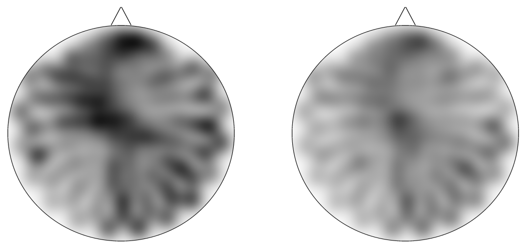
(3) いくつかの種類の最小二乗確率的分類法を統一的な形で再定式化を行った。その新しい定式化のもと，カテゴリ推定の信頼度を直接的に求めることが可能という特徴を残したまま，マルチラベル分類に対応できる新しいアルゴリズムを開発した。また，確率密度関数間の L2 距離を直接推定できる画期的な手法を考案した。さらに，その収束性能やバイアスなどの数理的な性質を明らかにしたとともに，効率良い実装方法を与えた。そして，密度差推定手法に基づく変化検知を開発し，その有効性を計算機実験により示した。

(4) Stiefel 多様体における経験的標本平均を求めるために，多様体から接空間への lifting と接空間から多様体への retraction および，接空間上での平均操作を，Cayley 変換，orthographic および QR 分解や極分解で与えた。これらは，x_k $(k = 1, 2, \ldots, N)$ から Kolmogoroff-Nagumo 平均 μ_{KN} を与える式

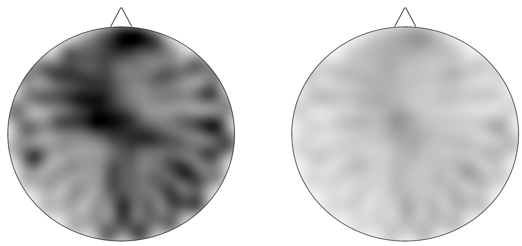
$$\mu_{KN} = \psi\left(\frac{1}{N}\sum_{k=1}^{N}\psi^{-1}(x_k)\right)$$

における ψ，および ψ^{-1} の役割を果す。しかしながら，これらの手法には，扱う行列のサイズが大きくなるにつれて，計算時間が飛躍的に増加するという問題が残された。この問題は，どの retraction/lifting 対を利用した平均演算も，retraction か lifting のどちらかに閉じた形では解けないものが存在することに起因している。例えば，極分解を基にした retraction/lifting 対の場合，retraction は閉じた形で解くことができるのに対し，対応する lifting は連続時間代数 Riccati 方程式を解く必要がある。これに対して，計算時間の増大を抑えるために，新たな Kolmogoroff-Nagumo 平均を用い，ともに閉じた形で与えられる極分解 retraction と orthographic lifting を組合わせ，高速に平均演算を実現できる手法を開発した。そして，開発した手法の有効性を，解析的，および数値的な側面から確認した。

(5) 適応カーネル主成分分析法を開発し，適応的かつ計算量を抑えた基底を選択するアルゴリズムを導入し，基底の数を抑えつつ，効率的な特徴抽出を可能にした。提案したアルゴリズムを物体追跡問題や，脳磁図からの特徴抽出に応用し，その性能を示した．脳磁図に応用した例を図 1 に示す。



(a) 原データ (閉眼時)　(b) 原データ (開眼時)



(c) 提案法 (閉眼時)　(d) 提案法 (開眼時)

図 1 α 波の分布図

この図は，開眼時と閉眼時の脳磁図信号に 8-12Hz の帯域通過型フィルタを適用することで α 波を抽出し，その大きさを濃淡で示したものである。原データでは，背景雑音の影響により閉眼時と開眼時での濃淡差が小さいが，提案法を適用した場合には，背

景雑音の影響が抑えられることにより濃淡差が大きくなり，開眼による α-block の効果がより強く確認できる。

５．主な発表論文等

[雑誌論文]（計　２８　件）

① S. Fiori, T. Kaneko, and T. Tanaka, Mixed maps for learning a Kolmogoroff-Nagumo-type average element on the compact Stiefel manifold, Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing, in press, 2014, 査読有

② T. Wakahara and Y. Yamashita, "k-NN classification of handwritten characters via accelerated GAT correlation," Pattern Recognition Vol. 47, No. 3, pp.994-1001, 2014, 査読有

③ Y. Urakami, K. Kawamura, Y. Washizawa, and A. Cichocki, Electroencephalographic gamma-band activity and music perception in musicians and non-musicians, Activitas Nervosa Superior Rediviva, Vol. 44, No. 4, pp.149–157, 2013, 査読有

④ T. Kaneko, S. Fiori, and T. Tanaka, Empirical arithmetic averaging over the compact Stiefel manifold, IEEE Transaction on Signal Processing, Vol. 61, No.4, pp.883–894, 2013, 査読有

⑤ H. Nam, H. Hachiya, and M. Sugiyama, Computationally efficient multi-label classification by least-squares probabilistic classifiers, IEICE Transactions on Information and Systems, Vol. E96-D, No. 8, pp.1871–1874, 2013, 査読有

⑥ M. Sugiyama, T. Suzuki, T. Kanamori, T., M. C. du Plessis, S. Liu, and I. Takeuchi, Density-difference estimation, Neural Computation, Vol. 25, No. 10, pp.2734–2775, 2013, 査読有

⑦ T. Yokota and Y. Yamashita, A Quadratically Constrained MAP Classifier Using the Mixture of Gaussians Models as a Weight Function, IEEE Trans. on Neural Networks and Learning Systems, Vol. 24, No. 7, pp.1127-1140, 2013, 査読有

⑧ N. Inoue and Y. Yamashita, Simultaneous learning of localized multiple kernels and classifier with weighted regularization, 2012 Proceedings of Joint IAPR International Workshops on Structural and Syntactic Pattern Recognition and Statistical Techniques in Pattern Recognition, pp.354–362, 2012, 査読有

⑨ Y. Washizawa, Adaptive subset kernel principal component analysis for time-varying patterns, IEEE Trans. Neural Networks and Learning Systems, Vol. 23, No.2, pp.961–1973, 2012, 査読有

⑩ Nopriadi and Y. Yamashita, A new approach to a maximum a posteriori-based kernel classification method, Neural Networks, Vol. 33, pp.247-256, 2012, 査読有

[学会発表]（計　１　件）

① 横田達也, 若原徹, 坂野鋭, 山下幸彦，正規分布に基づく Fisher 判別分析の補正項, パターン認識・メディア理解研究会, 2013.03.1, 東京

[図書]（計　１　件）

① 杉山将, イラストで学ぶ機械学習：最小二乗法による識別モデル学習を中心に, 講談社, 2013, 230pages

６．研究組織

(1) 研究代表者

山下　幸彦（YAMASHITA YUKIHIKO）
東京工業大学・大学院理工学研究科・准教授
研究者番号：９０２２０３５０

(2) 研究分担者

杉山　将（SUGIYAMA MASASHI）
東京工業大学・大学院情報理工学研究科・准教授
研究者番号：９０３３４５１５

田中　聡久（TANAKA TOSHIHISA）
東京農工大学・大学院工学研究院・准教授
研究者番号：７０３６０５８４

鷲沢　嘉一（WASHIZAWA YOSHIKAZU）
電気通信大学・大学院情報理工学研究科・助教
研究者番号：１０４１９８８０