

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 1 日現在

機関番号：12612

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24500165

研究課題名(和文) 広い分布族に対応するパラメトリック生成学習モデルと応用

研究課題名(英文) Generative model in a wide class of distribution and its application

研究代表者

高橋 治久 (Takahashi, Haruhisa)

電気通信大学・情報理工学(系)研究科・教授

研究者番号：90135418

交付決定額(研究期間全体)：(直接経費) 4,100,000円

研究成果の概要(和文)：本研究では、確率モデルによる特徴抽出と簡単な識別モデルによる学習機械について研究した。確率モデルとしてカーネルによりポテンシャル関数が定義されたマルコフ確率場，カーネルランダムフィールドを提案し，その効率的な計算法と特徴が線形分離可能となることを示した。最大の成果は，この特徴抽出器は，カーネルの次数を十分大きくとれば，パターン識別問題が必ず，識別の簡単な特徴である線形分離可能な問題に変換できることを示した点にある。顔検出問題にこの手法を用い，AdaBoostとSURFと用いた最先端の従来法の精度を容易に達成できることを示した。

研究成果の概要(英文)：We studied learning machine using the generative models and simple discriminator. We proposed a new framework of Markov random field that has kernelized potential function. We showed an efficient method of computation and that this model generates essentially linearly separable kernel features if the degree of kernel is large.

We conducted experiments on the face detection using the appearance based approach, and showed that our method can attain comparable results with the state-of-the-art face detection methods based on AdaBoost, SURF, and cascade despite of smaller data size and no preprocessing.

研究分野：情報通信(機械学習とパターン認識)

キーワード：自己相関カーネル サポートベクトルマシン Fisherカーネル 顔検出

1. 研究開始当初の背景

機械学習においては、深層学習などが提案されているが、学習規模が大きいため使い勝手は良くない。一般に、事例の分布に適合したモデルでなければ訓練事例での線形分離が必ずしも良い汎化性能を与えるとは限らない。深層学習、特徴の正規化によりこの事例の分布をうまく調整していると言える。SVM (サポートベクトルマシン) は、使い勝手がよいが、カーネルが事例の分布を反映している必用がある。しかし、このようなカーネルを得ることは、一般には難しい。Fisher カーネルはこのような流れのなかで提案されたものであるが、前提となる生成モデルの選択や計算の点から一般的なカーネルとはなっていない。SVM はディープラーニングに比べて、識別性能では敵わないがアルゴリズムの汎用性の点では優れている。ディープラーニングの汎化能力を備え、SVM のような汎用的アルゴリズムに近いものは理想と言える。カーネルによる特徴表現は、層を重ねなくても線形分離可能な特徴を得ることか出来るため、Fisher カーネル以外にも確率モデルとの組み合わせがいくつか考えられてきている。いずれも計算量の点で応用が難しい。

2. 研究の目的

識別学習のための良い特徴量は、その特徴の上でデータが線形分離可能となることが要求される。近年のディープラーニングの成功も特徴抽出を階層的に重ねることで、最終的に良い特徴量が得られることに起因する。言い換えれば、良い特徴量には、最終的な特徴空間上に写像されたデータが乗る多様体が、線形分離されることが要求される。本研究では、大局的に見て、確率モデルでの特徴抽出と線形 SVM のような簡単な識別機による識別という手法を提案する。とくに MRF (マルコフ確率場) の対数ホテンシャル関数と密接に関係している高次自己相関カーネルを提案し、多項式カーネルと同じ計算量のアルゴリズムを示す。更に、MRF の Fisher 特徴量のデータ多様体はカーネルの次数を十分取れば線形分離されることから、線形 SVM との組み合わせで非常に高性能な識別機が得られることを示す。

3. 研究の方法

本研究では、まず高次自己相関カーネルを提案する。このカーネルは MRF の対数ホテンシャル関数と密接に関係しており、多項式カーネルと同じ計算量のアルゴリズムを示す。このカーネルにより、多項式やカウスカーネルなどの SVM に比べ識別性能が格段に改善することを、顔検出による計算機実験で示す。また state-of-the-art な顔検出手法と同等の結果が得られる事を示す。なお、MRF の Fisher 特徴量を用いれば、データ多様体が線形分離される表現が得られることも確認する。こうして確立される識別機を種々のパタ

ーン認識へ応用する。

4. 研究成果

入力特徴を $x = (x_1, \dots, x_n)^t$ とする。このとき n 次までの自己相関ベクトルを

$$\phi(x) = (1, x_1, \dots, x_n; x_1x_2, x_1x_3, \dots, x_{n-1}x_n; \dots; x_1x_2 \cdots x_k, \dots, x_{n-k+1} \cdots x_n)^t$$

とすれば自己相関カーネルは、 $K(x, z) = \phi(x)^t \cdot \phi(z)$ で定義される。この自己相関カーネル $K(x, z)$ を用いてカーネル確率場(KRF)は以下のように定義される

$$P(x|\mu) = \frac{1}{Z} \exp\left(-\sum_{\ell=1}^m \mu_{\ell} K(\xi_{\ell}, x)\right)$$

但し、 Z はパーティションファンクション、 x は有限離散状態、 ξ はカーネル中心の事例、 μ はモデルパラメタである。

このモデルの事例 に対する対数尤度は

$$L_{\ell}(\mu) = \log P_{\mu}(\xi_{\ell})$$

で表され、このときフィッシャースコア

$$\frac{\partial L_{\ell}(\mu)}{\partial \mu_{\ell}} = K(\xi_{\ell}, \xi_{\ell}) - K(\xi_{\ell}, \bar{x})$$

は、カーネル事例数の次元の特徴量となる。この特徴量は、モデルと同じクラスの事例の場合には値は 0 に近く、そうでないときは大きくなる。言い換えれば、この特徴量はモデルのクラスと異なるクラスを、特徴量の成分が大きな要素が多い場合とそうでないもので識別する多数決問題と同じであり、線形分離可能な特徴量となっていることが分かる。

本研究の成果は、このフィッシャー特徴を線形 SVM で分離することにより、上に述べたように、理論的にパーフェクト汎化を与える学習識別器が構成できることであり、実験においてそれを確認した。

始めに、計算量の観点から、本手法の有効性を説明する。計算量は、自己相関カーネルの計算、KRF の計算の二項目からなる。以下順を追って説明する。

自己相関カーネルの計算

高次自己相関カーネルは、変数 x_1, \dots, x_n の基本対称式で表すことが出来る。基本対称式は以下のように定義される。

$$\begin{aligned} S_0(x) &= 1 \\ S_1(x) &= x_1 + \dots + x_n \\ S_2(x) &= x_1x_2 + x_1x_3 + \dots + x_{n-1}x_n \\ S_3(x) &= x_1x_2x_3 + x_1x_2x_4 + \dots + x_{n-2}x_{n-1}x_n \\ &\dots\dots\dots \\ S_n(x) &= x_1x_2 \cdots x_n \end{aligned}$$

基本対称式から、d 次の高次相関カーネルを作るために、 $y_i = x_i z_i$ とすれば、上式の d 次までの内積は

$$K(x, z) = \sum_{i=0}^d S_i(y), \quad y = (y_1, \dots, y_n)^t$$

となり、計算アルゴリズムは学会発表文献の補題一に基づいて、 d を固定したときに線形オーダーで与えられる。実際、図 1 に示すような線形の計算時間となっている。

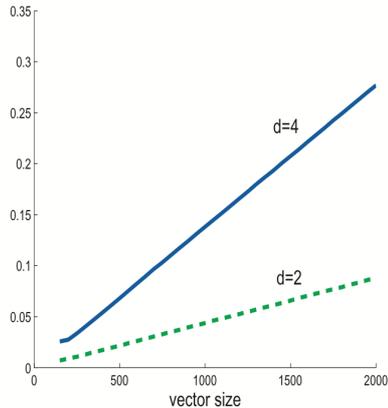


図 1 自己相関カーネルの計算時間

KRF の計算

この計算は、計算量が指数オーダーになるため直接確率関数を求める事は困難である。そこで、平均場近似を用いることでこれを克服する。学会発表文献で示したように、平均場の周辺確率は次式で与えられる。

$$r_j^k = \frac{\exp\left(\sum_{\ell=1}^m k \xi_{\ell}^j K_{d-1}(\xi_{\ell}^{(j)}, \bar{x}^{(j)}) \mu_{\ell}\right)}{1 + \sum_{k'=1}^{r-1} \exp\left(\sum_{\ell=1}^m k' \xi_{\ell}^{(j)} K_{d-1}(\xi_{\ell}^{(j)}, \bar{x}^{(j)}) \mu_{\ell}\right)}$$

但し、カーネル記号の添え字の $d-1$ はカーネルの次数を表す。

以下では応用を示す。

(1) 顔検出

パターン認識実験では、顔検出は典型的な問題であり、実用上も大事である。顔検出は、Viola-Jones 等のブレイクスルー以降、SURF 特徴量と AdaBoost、カスケードによる学習により改良されてきている。合わせて、顔検出のデータベースも、更新されており、過去の結果と比べて純粋に機械学習の性能を比較評価するのは難しくなっている。提案手法の識別性能を評価するため、既に時代遅れとなっはいるが CMU+MIT データセットを用いた。このデータセットは、CMU+MIT データセットのサブセットであり、この分野の有名な結果である Viola-Jones 等の検出器も同じデータセットのテストデータを用いて行われている。データは 19×19 ピクセルサイズで、256 階調のグレースケールである。トレーニング用に、顔データが 2429 個、非顔データが 4548 個用意されている。またテストデータは顔が 472 個、非顔が 23573 個ある。実験では、トレーニング用に、顔データ 2429 個すべてを用い、非顔データは 2429 個を反転したデータを加えて、6644 個のデータを用意した。テストセットは、顔デ

ータ、非顔データ共に上記テストデータからランダムに抽出した 470 個を用いた。本アプローチでは、特徴抽出を行わず、上記グレースケールデータを直接入力とするアピアランススペースで行った。但し、入力を正規化するため、射影空間への変換を行った。これは、各画像をそのノルムで割り、調整のための係数 $19/2$ をかけている。こうすることで、ほとんどのピクセル値は 0 と 1 の間に収まる。図 2 に、カーネル数 1200 で、4 次自己相関カーネルを用いた識別結果を示す。参考のため KRF の学習回数 10 回で線形 SVM を用いた場合 ROC カーブを示す。最新の顔検出の分野では、Viola-Johns により提案された、'Integral features' と呼ばれる単純な特徴の集合で画像を現し、AdaBoost によって critical な visual features を取捨選択し、最後に識別器を cascade に動作させることで、非顔が確実であるようなデータをいち早く捨てることで、計算時間の短縮を狙った手法を基本としている。最近の結果では、豊富なデータセットで学習を行っているため、必ずしも識別器の性能の比較には成らないが、CMU+MIT テストセットを用いて Viola-Johns 等の結果との比較があるので、論文のグラフより、値を拾って、表 1 に比較を示し参考にする。この表では Viola-Johns 等の false detection 10% での検出率が分かっているので、対応する値をグラフから拾うことで、最新の結果を引用する。本手法が、豊富でないデータでの学習にもかかわらず、最新の結果と comparable であることが分かる。

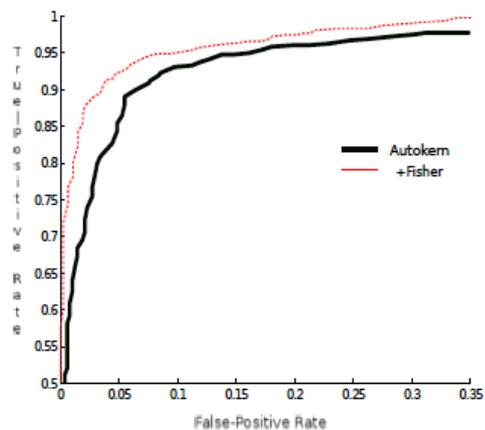


図 2 ROCカーブ

表 1: Recognition/detection rates(%) for false positives (10, 20%) on CMU+MIT test set

| Systems \ False positives | 10 % | 20% |
|--|------|------|
| Autocorrelation kernel on the subset of CMU+MIT test set | 93.0 | 96.0 |
| Viola and Jones([11]) | 92.1 | 93.2 |
| SURF cascade, J.Li and Y. Zhang (picked up from [8]) | (94) | - |

確率的な特長抽出を簡単な線形識別帰で

識別する本手法は、疑似確率を SVM で得ることによって人間の行動認識へ適用し成功した(雑誌論文)。

特徴量は ST-patch (2次元画像を時間方向に重ねた3次元データにおける局所領域内の時空間特徴領域内の画素値の勾配から計算される特徴)とオプティカルフローを用いている。オプティカルフローは強度が強い順にソートし、上から N 個取得する。こうすることにより、人の動作とは関係のない小さな動きを除去することが出来る。次に、ノイズに対して更に強くなるようも行った。

識別は、一段目に SVM を用いた疑似確率による疑似生成モデルで特徴抽出を行い、二段目で角自国語との識別を SVM で行っている(図3)。

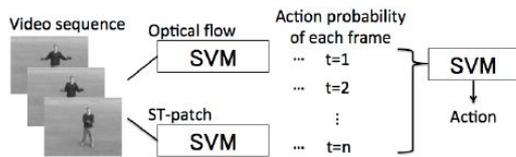


図3 行動認識の識別機の構成

提案手法の評価に、KTH データセットを使用した。このデータセットは、6種類の動作(腕の動作である boxing, handclapping, handwaving、移動する動作である jogging, run-ning, walking)が25人分、2391シーケンス存在する。SVMによる学習は24人分の動画で行い、残りの1人をテスト用に用いる leave-one-out 方式で実験を行い、25人分の識別を行った。

表2は手法ごとの識別率を示す。平均で識別率は95.2%となり、最高水準の結果を簡単な方法で得ている。

表2: 動作ごとの識別結果(提案手法:two-level SVMs)

| Action | box | clap | wave | jog | run | walk |
|--------|------|------|------|------|------|------|
| box | 95.0 | 2.5 | 1.3 | 0.0 | 0.3 | 1.0 |
| clap | 2.7 | 96.5 | 0.8 | 0.0 | 0.0 | 0.0 |
| wave | 1.3 | 0.3 | 98.5 | 0.0 | 0.0 | 0.0 |
| jog | 0.0 | 0.0 | 0.0 | 90.0 | 9.8 | 0.3 |
| run | 0.0 | 0.0 | 0.0 | 7.3 | 92.8 | 0.0 |
| walk | 0.0 | 0.0 | 0.0 | 0.7 | 0.5 | 98.8 |

カーネルパラメタを最適に設定することは、重要な問題である。雑誌論文において、カーネルPCA(主成分分析)の観点からこの方法を開発し、テストデータでその効果を確認できた。

KSC(Kernel Spectral Clustering)はスペクトラルクラスタリングをカーネルPCAと組み合わせる事によって精度向上を実現したものである。その性能はカーネルパラメタに強く依存する。最適パラメタ決定手法は網羅的探索が基本であり、従来の経験的手法を

脱したとは言い難い状況である。

本研究では KSC のカーネルパラメタを自動決定する手法のパラメタ学習手法とグラフラプリアン表現を用いることで実現した。この学習手法はグラフラプリアンの固有値特徴を最適化するパラメタを発見するので、真に定性的な最適化が行える。従来、実現事例のない問題として、動画サイトのタグに着目し動画のカテゴリ分類を行った。k-means 法と比較を行い、本手法が十分な性能が良いことを計算機実験により示した。

KSC は SVM と重み付きカーネル PCA によって特徴空間上で分散最大の射影を以下の最適化問題を解くことで構成する

$$\min_{w^{(l)}, e^{(l)}, b_l} \frac{1}{2} \sum_{l=1}^{n_e} w^{(l)\top} w^{(l)} - \frac{1}{2N} \sum_{l=1}^{n_e} \gamma_l e^{(l)\top} V e^{(l)}$$

$$\text{subject to: } e^{(l)} = \Phi w^{(l)} + b_l \mathbf{1}_N$$

ここで $w^{(l)}$ は係数ベクトル、 1 は正規化定数、 V は射影行列、 b_l はバイアス、 V は重み付け行列である。この問題は双対問題に変えて解かれる。双対問題中ではスペクトラルクラスタリングを行っているので KSC ではスペクトラルクラスタリングにおける完全連結な重み関数としてカーネルパラメタを伴うガウス基底関数形のカーネル関数を選択する必要がある。入力データ x_i と x_j に対するカーネル関数は

$$K(x_i, x_j) = \exp \left[- \left(\frac{d(x_i, x_j)}{\sigma} \right)^2 \right]$$

このカーネルから得られるグラム行列と重み行列 V から得られるグラフラプリアン行列の固有値スペクトルは(パラメタの設定により)次の性質を満たせば最適である。

連結成分(クラスタ)の数だけの固有値が0に近いことが最適な固有値特徴として求められる。KSC においては $k-1$ 個の連結成分を構成することが求められることより、全固有値を昇順に並び替えた時、 $k-1$ 番目の固有値と k 番目の固有値差が1に近い事が最適状態であるといえる(図4参照)。

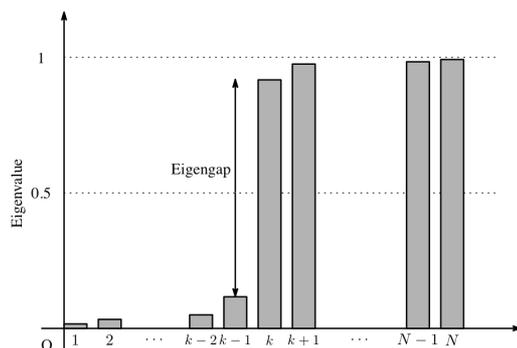


図4 固有値スペクトル

この条件を得るために、以下のように評価関数を定め、固有値の導関数を導き、最適値を

求めた。

$$E(\theta) = \frac{1}{(k-1)^2} \left[\sum_{i=1}^{2k-2} \mu_i - (k-1) \right]^2$$

$$\frac{\partial E(\theta)}{\partial \theta} = \frac{2}{(k-1)^2} \left(\sum_{i=1}^{2k-2} \frac{\partial \mu_i}{\partial \theta} \right) \left[\sum_{i=1}^{2k-2} \mu_i - (k-1) \right]$$

この方法を、単純な比較実験として、動画サイト 2 の投稿動画に付与されたタグの発生情報を元にベクトル空間モデルを生成し、主要カテゴリ (ゲーム、アニメ、音楽、政治、スポーツ、科学) から 2 カテゴリを選び出し 2 クラス分類実験を行い k-means 法と純度平均を比較した結果を表 3 に示す。

表 3 KSC vs k-means

| KSC \ k-means | 1 | 2 | 3 | 4 | 5 | 6 |
|---------------|------|------|------|------|------|------|
| ゲーム (1) | - | 80.7 | 79.2 | 80.0 | 80.4 | 73.9 |
| アニメ (2) | 95.8 | - | 79.0 | 81.1 | 78.7 | 79.1 |
| 音楽 (3) | 94.3 | 95.1 | - | 82.5 | 75.9 | 78.7 |
| 政治 (4) | 97.9 | 96.7 | 95.2 | - | 76.7 | 78.0 |
| スポーツ (5) | 84.7 | 96.7 | 91.1 | 90.8 | - | 78.0 |
| 科学 (6) | 99.9 | 97.3 | 94.0 | 92.1 | 92.4 | - |

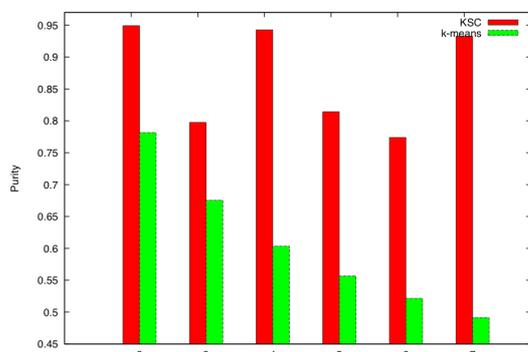


図 5 カテゴリ数が増大したときの違い

更に 2 カテゴリから 7 カテゴリにおける純度の比較は図 5 の様になった.KSC ではカテゴリ数が増加しても分類率が低下せず、高い純度を保持した。

以上より最適パラメタ設定法は有効であることが確認できた。

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 2 件)

Taiyo Mineo and Haruhisa Takahashi, Automatic Kernel Parameter Tuning of KSC for Video Category Classification, Journal of Signal Processing, 18(04) 237-240, Jul. 2014

Manabu Yoshida and Haruhisa Takahashi, Human Action Recognition with Two-Level SVMs, Journal of Signal Processing, Vol.17, No.4, pp.159-162, July

2013

〔学会発表〕(計 13 件)

Sho Torii, Haruhisa Takahashi, LatentSVM using 2 kernel, Proceedings of 2015 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing, 534-537, Kuala Lumpur, Malaysia, February 27 – March 2, 2015

Masahiko Tsujita, Haruhisa Takahashi, Human Action Recognition using Deep Recurrent CNN Proceedings of 2015 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing, 533, Kuala Lumpur, Malaysia, February 27 – March 2, 2015(アブストラクト)

Kazuki Kubota, Haruhisa Takahashi, Human Action Recognition Using MRF and HOG Features Proceedings of 2015 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing, 538, Kuala Lumpur, Malaysia, February 27 – March 2, 2015(アブストラクト)

Yasuo Shimada, Haruhisa Takahashi, Human Action Recognition using Deep Recurrent CNN, Proceedings of 2014 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing, 167-174, Honolulu, Hawaii, USA, February 28 – March 3, 2014

Kohei Egami, Haruhisa Takahashi, Single Image Depth Estimation From SVR and DRF, Proceedings of 2014 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing, 373-376, Honolulu, Hawaii, USA, February 28 – March 3, 2014 (受賞)

Taiyo Mineo, Haruhisa Takahashi, Auto Kernel Workshop on Nonlinear Circuits, Communications and Signal Processing, 273-276, Honolulu, Hawaii, USA, February 28 – March 3, 2014 (受賞)

Haruhisa Takahashi, Learning with Kernel Random Field and Linear SVM, Proceedings of the, 3rd International Conference on Pattern Recognition Applications and Methods (ICPRAM2014), ESEO, Angers, Loire Valley, France, pp 167-174, 6-8, March 2014(査読あり、高評価)

高橋治久：自己相関カーネルによる学習，第 36 回情報理論とその応用シンポジウム，11 月 26-29,伊東ホテル聚楽，伊東市，2013.

Sadaaki Yamazaki, Haruhisa Takahashi, Image Labeling with Conditional Random Fields using Mean Field Approximation, International Workshop on Nonlinear Circuits, Communications and Signal Processing, Courtyard King Kamehameha'

s Kona Beach Hotel, The Isrand of Hawaii, Hawaii, USA, 369-372, March 4 – 7, 2013.

Akihiro Yamamoto , Haruhisa Takahashi, Image Categrization using HDP with KPCA for Feature Transformation, International Workshop on Nonlinear Circuits, Communications and Signal Processing, Courtyard King Kamehameha ' s Kona Beach Hotel, The Isrand of Hawaii, Hawaii, USA, 381 -384, March 4 – 7, 2013.

Masaaki Sugiyama, Haruhisa Takahashi, Improving Semi-Supervised Learning for Big Image Collections ,International Workshop on Nonlinear Circuits, Communications and Signal Processing, Courtyard King Kamehameha ' s Kona Beach Hotel, The Isrand of Hawaii, Hawaii, USA, 540-543, March 4 – 7, 2013.

Manabu Yoshida , Haruhisa Takahashi, Human Action Recognition with Two-Level SVMs , International Workshop on Nonlinear Circuits, Communications and Signal Processing, Courtyard King Kamehameha ' s Kona Beach Hotel, The Isrand of Hawaii, Hawaii, USA, 580-583, March 4 – 7, 2013. (受賞)

Kouhei Egami, Haruhisa Takahashi, Depth Estimation with SVR and CRF , International Workshop on Nonlinear Circuits, Communications and Signal Processing, Courtyard King Kamehameha ' s Kona Beach Hotel, The Isrand of Hawaii, Hawaii, USA, 657-660, March 4 – 7, 2013.

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

〔その他〕

ホームページ等

6 . 研究組織

(1)研究代表者

高橋 治久 (TAKAHASHI HARUHISA)
電気通信大学・大学院情報理工学研究科・
教授
研究者番号：90135418

(2)研究分担者

堀田 一弘 (KAZUHIRO HOTTA)
名城大学・理工学部・准教授
研究者番号：40345426