

## 推薦論文

# 行動認識ニューラルネットワークによる特徴抽出の可視化と転移学習への応用

吉村 直也<sup>1,a)</sup> 前川 卓也<sup>1,b)</sup> 原 隆浩<sup>1,c)</sup>

受付日 2020年5月7日, 採録日 2021年2月2日

**概要:** 行動認識分野においてもニューラルネットワークを用いた転移学習が注目を集めている。ニューラルネットワークの学習には大量のデータが必要であるが、転移学習を利用することで、既存の大規模データセットを有効活用し、学習データの収集コストを削減することができる。しかし、行動認識データセットは対象とする行動セットが異なるなどの問題のために、想定する利用環境と同じ想定で収集されたデータセットを見つけることは難しく、容易に転移学習を利用することができない。そこで、本研究は転移学習の有効活用に向けて、行動認識ニューラルネットワークが抽出する特徴の可視化と、それを用いた転移元データセットの推薦手法を提案する。本研究ではまず、画像分野の可視化手法である Activation Maximization (AM) を行動認識モデルに適用することを試みる。画像信号を想定して開発された AM を、行動認識モデルが扱う加速度信号にそのまま適用すると、ノイズや異常値を多く含む信号が生成される。本研究ではこれを抑制するため、より行動認識モデルに適した正則化手法を提案する。また提案する可視化手法の応用例の1つとして、想定する環境に適した転移学習のための転移元データセットを推薦する手法を提案し、6種類のデータセットを用いてその有効性を確認した。

**キーワード:** 行動認識, ニューラルネットワーク, 可視化技術, 転移学習

## Visualization of Deep Feature Representation Toward Transfer Learning of Activity Recognition Models

NAOYA YOSHIMURA<sup>1,a)</sup> TAKUYA MAEKAWA<sup>1,b)</sup> TAKAHIRO HARA<sup>1,c)</sup>

Received: May 7, 2020, Accepted: February 2, 2021

**Abstract:** Transfer learning using neural networks has been attracting attention in the activity recognition research field. Transfer learning can reduce the amount of required training data by leveraging existing large datasets. However, it is difficult to find a dataset collected under the same sensor conditions as the target environment due to problems such as the difference in assumed activity classes, making it difficult to apply transfer learning. In this study, we propose a method to visualize/generate features extracted by a activity recognition neural network and to recommend a source dataset based on the visualization technique. In this study, we first try to apply Activation Maximization (AM), a visualization method for images, to activity recognition models. When the AM methods developed for the image signal is applied to acceleration signal, which is the input of the activity recognition model, the generated signals contain noises and outliers. In this study, we propose a regularization method that is more tailored activity recognition models in order to suppress this problem. We also applied the proposed visualization method to recommend a source dataset for transfer learning that is suitable for a given environment, and confirmed the effectiveness of the method using six datasets.

**Keywords:** human activity recognition, neural network, visualization, transfer learning

<sup>1</sup> 大阪大学大学院情報科学研究科  
Graduate School of Information Science and Technology,  
Osaka University, Suita, Osaka 565-0871, Japan

a) yoshimura.naoya@ist.osaka-u.ac.jp

b) maekawa@ist.osaka-u.ac.jp

c) hara@ist.osaka-u.ac.jp

本論文の内容は2019年7月のマルチメディア、分散、協調とモバイル (DICOMO2019) シンポジウムで報告され、ユビキタスコンピューティングシステム研究会主催により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である。

## 1. 概要

行動認識分野においてもニューラルネットワークを利用した研究が多く行われている。行動認識技術はウェアラブルデバイスから取得した加速度やジャイロ信号などのセンサデータからユーザの行動の内容を推定する技術である。ニューラルネットワークを行動認識に用いることで、(1) 高い精度での認識と、(2) 特徴抽出の自動化という利点があるが、その一方で学習には非常に大量のデータを必要とするという欠点がある。この問題を解決するために転移学習が注目されている。転移学習は、大規模データセットを用いて事前に学習したモデルを活用することで、転移先環境において必要な学習データ量を削減できる。

しかし、行動認識分野において転移学習の利用は容易ではない。データセットによって行動クラスやセンサの装着位置など想定する環境が異なるため、想定する環境に合うデータセットを見つけることは難しい。そのため研究者は想定する環境のデータを独自に大量に収集する必要がある。行動認識分野において転移学習を有効に活用できれば、少量のデータでも認識手法の検証が可能になり、行動認識技術の発展を加速させることができる。そこで本研究では、行動認識分野における転移学習の有効活用に向けた、行動認識モデルの分析と転移学習を行うための転移元データセットを推薦する技術を提案する。

本研究では、行動認識ニューラルネットワークの分析のため、ネットワークが抽出する特徴量を明らかにする。本研究では、CNN を用いたニューラルネットワークを対象とし、画像処理分野の可視化手法である「Activation Maximization (AM)」[5] を行動認識モデルに適用する。AM は、ネットワーク中の注目したユニットの出力値を最大化する入力信号を生成することで、ユニットが抽出する特徴を入力信号として表現する手法である。本研究は行動認識モデルの入力として加速度信号を想定するが、AM は画像信号を想定した手法である。加速度信号は、値域が 0–255 に限定されている画像信号に比べて、裾の長い分布をしている。そのため、入力信号を生成する過程で、異常に大きな振幅を持つ信号など、ノイズを多く含む現実の加速度信号とは大きく異なる信号が生成される傾向がある。本研究ではこの問題を解決するために、このような異常値を抑制する行動認識モデルに適した正則化手法を提案する。

次に可視化手法の応用例の 1 つとして、転移元データセットの推薦を行う。本研究では、転移元データではなく、それを学習したモデルに注目し、転移元モデルが抽出する特徴が転移先にどの程度適合するかを求めることで、転移学習後の認識精度が高くなると期待されるデータセットを推薦する。一般に転移元のデータセットを選択する場合は、データセットに含まれる行動クラスなどを参考にするが、データセットに含まれる行動クラスが一致しない場合は多い。

一方で、転移先のデータセットに含まれる行動クラスを認識するために必要な特徴を学習できるデータセットは数多く存在する。たとえば、転移先データセットに「Using PC」という行動クラスがある場合、このクラスの認識に必要な手の動きに関する特徴は「Writing Notebook」などの行動クラスからも学習できると考えられる。本研究では、提案する可視化手法を用いることで、転移元データセットで学習できる特徴に基づいて、転移元データセットの推薦を試みる。本研究の学術的貢献点は以下のとおりである。

- 行動認識モデルに対して Activation Maximization を適用し、ネットワーク中の注目したユニットが抽出する特徴の可視化を行う。行動認識のデータに適した可視化を行うために、異常値の発生を抑制する正則化手法を提案する。
- 行動認識モデルの可視化手法を応用し、学習される特徴の傾向をデータセット間で比較した。またこの分析結果に基づいて、転移学習の転移元データセットを推薦する手法を提案し、6 種のデータセットを用いて有効性を確認した。

まず 2 章において、ネットワークの可視化手法および行動認識の転移学習に関する研究を紹介する。3 章では本研究で用いる可視化手法の詳細を説明し、4 章で評価を行う。最後に、5 章において実際に転移学習における転移元データセットの推薦手法を提案し、評価を行う。

## 2. 関連研究

### 2.1 ニューラルネットワークの可視化

ニューラルネットワークの検証や改良のためには、ブラックボックスとされるネットワークの内部処理を明らかにすることが重要である。画像処理分野では特に可視化技術が研究されており、モデルの判断根拠を可視化する手法 [18] や、各層で抽出される特徴マップや演算内容を可視化する手法が提案されている [9]。

本研究では Activation Maximization (AM) と呼ばれる、ネットワーク中の注目したユニットが最も強く反応する入力信号を生成する手法を用いる。先に紹介した手法は出力層や層の粒度での可視化を行うが、AM はさらに詳細なユニット単位での可視化を可能とする。AM を出力層や中間層に適用することで、ネットワークが獲得した特定の行動クラスのコンセプトや、そのユニットの機能を可視化することができる [7], [11], [12], [20]。先行研究では、AM が生成する信号には高周波成分などのノイズが多く含まれることが知られており [20]、これらのノイズを抑制するために L2 ノルムや Total Variation [16] \*1 など、正則化項の導入が検討されている。本研究では行動認識モデルに適した正則化項を提案する。

\*1 Total Variation は画像処理分野で用いられる正則化項であり、隣接する要素の差分を計算することでスムージングを行う。

## 2.2 行動認識における転移学習

行動認識においても転移学習の利用が活発に研究されている。Wang ら [19] による Maximum Mean Discrepancy (MMD) を用いた転移手法や, Khan ら [6] による Domain Adaptation に基づいた転移学習手法が提案されている。

行動認識では, 行動セットやセンサの装着方法などに大量の組合せがあり, 転移先の想定と合致するデータセットを探すことは難しい。したがって, 少し想定が異なっていたとしても, できるだけ転移先のデータに類似した転移元データを選択して使用する必要がある。Wang ら [19] は, 転移元と転移先の識別性能を元にドメイン間の距離を推定することで, 転移元データセットを推薦する手法を提案している。しかし, この手法は候補となるデータセットに研究者が主観的に重みを与えるなど, 完全に自動化されていない部分がある。本研究では, 可視化結果を用いて学習済みモデルが抽出する特徴表現の観点から, 自動的に転移元に適したデータセットを推薦する手法を提案する。

## 3. 行動認識ネットワークが抽出する特徴表現の可視化

本章では行動認識ニューラルネットが抽出する特徴表現を可視化する手法に関して説明する。まず, Activation Maximization に関して説明する。次に可視化結果に含まれるノイズを抑制するため, 行動認識モデルに適した正則化項を検討する。

### 3.1 本研究の想定

本研究では, ユーザの身体に装着された 3 軸加速度センサのセンサデータを入力とし, 事前に学習された CNN ベースの行動認識モデルに対する分析を想定する。行動認識モデルの入力  $\mathbf{X}$  は, 長さ  $N_T$  で切り出したウィンドウである。また,  $l$  層目のユニット  $u$  が出力する特徴マップを,  $\mathbf{X}^{(l,u)}$  とする。 $\mathbf{X}$  と  $\mathbf{X}^{(l,u)}$  は, それぞれ  $(N_S \times N_T)$  と  $(N'_S \times N'_T)$  の 2 次元のテンソルである。 $N_S$  は入力のセンサ数<sup>\*2</sup>,  $N'_S$  と  $N'_T$  は畳み込みフィルタなどの設定によって変わる定数である。

### 3.2 Activation Maximization

Activation Maximization (AM) [5] では, ユニットの出力値はそのユニットが抽出する特徴表現の強さを反映していると考えられる。AM は注目したユニットに対して最も大きな出力値を出す入力信号を勾配法によって生成することで, そのユニットが抽出する特徴量を, 入力信号として可視化する。可視化対象のモデルにおいて,  $\mathbf{X}$  を入力とし,  $l$  層目のユニット  $u$  の出力値を計算する関数を  $f^{(l,u)}(\cdot)$  と

すると, AM の目的関数は以下の式で表現できる。

$$\mathbf{X}^* = \arg \max_{\mathbf{X}} (f^{(l,u)}(\mathbf{X}) - wR(\mathbf{X})) \quad (1)$$

$\mathbf{X}$  は生成している入力信号,  $\mathbf{X}^*$  は目的のユニット  $u$  の出力を最大にする入力信号,  $R(\cdot)$  は後述する正則化項である。 $w$  は正則化項の強さを制御する重みである。また, 注目するユニットが出力する特徴マップ  $\mathbf{X}^{(l,u)}$  は 2 次元のテンソルであるが, ここでは  $f^{(l,u)}(\mathbf{X})$  によって要素の平均を取ることでスカラ値に変換している。つまり,  $f^{(l,u)}(\mathbf{X}) = \text{Mean}(\mathbf{X}^{(l,u)})$  である。本研究では畳み込み層を可視化対象とする。

式 (1) は勾配法によって解くことができる。具体的には, (1) まず入力信号  $\mathbf{X}$  を適当な値で初期化する。(2) 次に  $\mathbf{X}$  をモデルに入力し, 注目したユニットの出力値  $f^{(l,u)}(\mathbf{X})$  を得る。(3) 以下の式 (2) によって  $\mathbf{X}$  を更新する。

$$\mathbf{X} \leftarrow \mathbf{X} + \eta \frac{\partial}{\partial \mathbf{X}} (f^{(l,u)}(\mathbf{X}) - wR(\mathbf{X})) \quad (2)$$

$\eta$  は学習係数である。この (1)–(3) の手順を収束するまで繰り返すことで, 目的の信号  $\mathbf{X}^*$  が得られる。ニューラルネットワークの学習は勾配降下法を用いて, 損失関数を 0 にするようにモデルのパラメータを修正する。しかし, AM では式 (1) の右辺の括弧内を最大化するように入力信号のみを更新することで, 目的の信号  $\mathbf{X}^*$  が得られる。

上記の方法によって, 注目したユニットが最も強く反応する信号を生成することができるが, AM には次の問題が指摘されている。(問題 1) Yoshinski らの報告によると, 勾配法を用いて  $\mathbf{X}^*$  を生成すると高周波成分を多く含む波形が発生する [20]。(問題 2) ユニットの出力値を最大化する過程で, 異常に大きな振幅やユニットの出力値が生成されてしまう。このような可視化結果は人間が見ても理解ができない [13]。これらの問題を解決するために, 正則化項  $R(\mathbf{x})$  が重要な役割を果たす。先行研究では, L2 正則化や Total Variation, GAN を用いた正則化手法が検討されている [12], [20]。これらの正則化手法は画像信号に対して研究されているが, 加速度信号は画像信号と特性が異なる。たとえば, 画像信号は値域が 0–255 に制限されているが, 加速度信号はこれよりも裾の長い分布であり, 異常値の発生などに影響しやすい傾向にある。画像に対する正則化手法とは異なる, 加速度信号に適した正則化手法が必要である。

### 3.3 異常値を抑制するための正則化項

AM を用いることで, ユニットの出力値が異常に大きい信号が生成されてしまう。たとえば, 注目するユニットが 1 層目にあり, そのユニットの畳み込みフィルタの重みがすべて正かつ, 活性化関数が ReLU だとする。この場合, 入力値とユニットの出力値は比例関係にあるため, AM では振幅が大きくなるように入力信号を更新し続ける。入力

<sup>\*2</sup> 本研究では, 3 軸加速度センサの各軸を独立のセンサと見なす。したがって, 1 つの 3 軸加速度センサを用いる場合は  $N_S = 3$  となる。

信号の振幅に対して正則化を行うことでこれを抑制することが考えられる。しかし、前述のとおり加速度信号は裾の長い分布であり、ユニットごとに応答する値の範囲などが異なると考えられる。振幅を制限することだけでは異常値を抑制することは難しい。

そこで、本研究ではユニットの出力値に対する制約を与えることで異常値の発生を抑制する Extreme Activation Penalty (EAP) を提案する。ユニットの出力値が異常に大きな値をとる場合、対応する入力信号にも異常値が発生していると考えられる。したがって、このユニットの出力値に対してペナルティを与えることで、入力信号に現れる異常値も抑制することができる。異常値の検出は  $2\sigma$  法によって行う。個々のユニットの特性を考慮して入力信号に直接制約を与えることは難しいが、ユニットの出力値の分布は学習データをモデルに入力することで得ることができるため、それを用いて入力信号に間接的に制約を与えることができる。この分布からユニットごとに  $2\sigma$  法によって閾値  $th^{(l,u)}$  を決める\*3ことで、ユニットの特性を考慮した異常値判定を行う。以上を考慮して、EAP を以下の式 (3) で定義する。

$$R_{EAP}(\mathbf{X}^{(l,u)}) = \text{Mean}(\text{LeakyReLU}(\mathbf{X}^{(l,u)} - th^{(l,u)})) \quad (3)$$

EAP は基本的には、ユニットの出力値の大きさに比例したペナルティをかける。ただし、異常値とそれ未満の出力値で、 $R_{EAP}(\mathbf{X}^{(l,u)})$  への寄与の大きさを Leaky ReLU [8] を用いて変える。EAP ではまず、ユニットの出力値のテンソル  $\mathbf{X}^{(l,u)}$  の各要素から  $2\sigma$  法で決定した閾値  $th^{(l,u)}$  を引き、閾値を超えた出力値は正の値を、超えていないものは負の値を持つように変換する。これに対して、式 (4) の Leaky ReLU 関数を適用する。これによって閾値の前後で傾きを変え、 $R_{EAP}(\mathbf{X}^{(l,u)})$  への寄与の大きさを変える。閾値以下の値に適用される傾きは、 $\epsilon$  によってコントロールする。本研究では、 $\epsilon = 10^{-4}$  を用いており、閾値以下の値の寄与は限定的であるため、閾値を超えた異常値を重点的に抑制することができる。

$$\text{LeakyReLU}(x) = \begin{cases} x + \epsilon & (x \geq 0) \\ \frac{\epsilon}{th^{(l,u)}} \cdot x & (\text{otherwise}) \end{cases} \quad (4)$$

## 4. 可視化手法の評価と抽出される特徴の分析

本章では、前章で提案した手法を、6 種類の公開されている行動認識データセットに適用し評価を行う。

### 4.1 行動認識データセット

本研究では、以下のデータセットを用いる。

- **DSADS** : 「Daily and Sports Activities Data Set」 [3]

\*3  $2\sigma$  法では、分布の平均  $\mu^{(l,u)}$ ・分散  $\sigma^{(l,u)}$  を用いて、閾値は  $th^{(l,u)} = \mu^{(l,u)} + 2\sigma^{(l,u)}$  で定められる。

は、日常生活とスポーツにおける行動 19 種類に関して、8 人の被験者からデータを収集したものである。

- **MHEALTH** : 「MHEALTH Dataset Data Set」 [2] は、日常生活と運動に関する 12 種類の行動に関して、10 名のユーザからデータを収集したものである。
- **OPP-LOC/OPP-GES** : 「OPPORTUNITY Activity Recognition Data Set」 [15] は、日常生活における行動データを 4 名の被験者から収集したものである。行動ラベルは 2 種類あり、「Stand, Walk, Sit, Lie」の 4 種類の姿勢がラベリングされている「Locomotion (OPP-LOC)」と、「Open Door1」など屋内における 17 種類の行動がラベリングされた「Gesture (OPP-GES)」がある。OPP-GES は、本研究が想定する手と足のセンサだけでは認識が難しいため、同系統の行動を集約した 5 種類の行動を認識する。
- **PAMAP2** : 「PAMAP2 Dataset」 [14] は、日常生活の行動とエクササイズの詳細データを 9 名の被験者から収集したものである。ユーザによって行動セットが若干異なっており、それらを除外した 12 種類の行動を本研究では使用する。
- **REALDISP** : 「REALDISP Activity Recognition Dataset」 [1] は、日常生活と運動に関して 33 種類の行動を、16 人の被験者から収集したものである。本研究ではデータセット作成者が選択した 20 種類の行動を認識する。

本研究ではこれらのデータセットから、右手と左足首に装着した 3 軸加速度センサのデータを用いる。本研究では、全体のおおよそ 2 割の被験者をテストデータとし、残りを学習データとして用いた。すべてのデータセット共通の前処理として、まずサンプリングレートを 30 Hz に揃えたうえで、 $[-3G, +3G]$  を  $[-1, 1]$  に変換する正規化を行った。次にウィンドウ幅 1 sec (30 サンプル) のスライディングウィンドウを適用した。また行動ラベルが付与されていないサンプルは除外した。

### 4.2 行動認識モデル

本研究で使用する行動認識モデルを図 1 に示す。Münzner ら [10] のモデルを参考にし、畳み込み層 3 層と全結合層 3 層の合計 6 層によって構成する。また、畳み込みフィルタのサイズを  $f \times 1$  とした。  $f$  は時間方向の長さである。この設定によってモデルは、センサ・軸ごとに同じ畳み込みフィルタを適用して、特徴抽出を行う。

本研究では  $f = 5$  を用いた。また、ネットワークの学習は Adam optimizer を用いてバッチサイズは 1,024、学習エポック数は 300 とした。学習係数は  $lr = 10^{-4}$  である。モデルの評価は各行動クラスに対する F 値を平均した F1-measure (macro average) を用いる。それぞれのモデルを学習した結果を表 1 に示す。

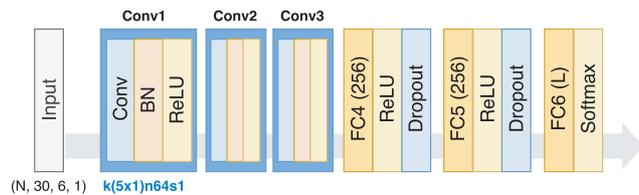


図 1 行動認識ネットワーク：Conv は畳み込み層（ユニット数は 64 で、ストライドは 1）、BN は Batch Normalization、FC は全結合層、L は行動クラス数を表す。入力合計 6 軸の加速度信号を 30 サンプル（= 1 sec）切り出したものとした

Fig. 1 Network architecture used in the experiment: “Conv”, “BN” and “FC” represent the convolutional layer, the batch normalization and the full-connected layer, respectively.

表 1 行動認識モデルの認識精度

Table 1 Recognition performance for each dataset.

Dataset	行動クラス数	F 値	学習に使用した被験者数
DSADS	19	0.741	6 人
MHEALTH	12	0.955	8 人
OPP-GES	5	0.477	4 人
OPP-LOC	4	0.891	4 人
PAMAP2	12	0.781	7 人
REALDISP	20	0.723	13 人

### 4.3 Activation Maximization の実行方法

基本的には 3.2 節に示した方法で AM を実行する。前述のモデルでは、全結合層の直前まではそれぞれのセンサが独立に処理され、センサごとに同じ特徴抽出の演算が行われる。そのため、各センサに同じ初期値を与えて、 $(N_S \times N_T) = (6 \times 30)$  の 2 次元テンソル  $\mathbf{X}$  に対して AM を実行すると、それぞれのセンサは同じ波形になる。本研究では、計算コストの削減のため任意のセンサ 1 つのみを対象として AM を行った。つまり、 $(N_S \times N_T) = (1 \times 30)$  の 2 次元テンソルに対して、AM を行った。また、本実験では最初に  $\mathbf{X}$  を、最大振幅 0.3 G、周波数 2 Hz の正弦波で初期化した。学習係数は  $\eta = 0.005$  とし、 $\mathbf{X}$  の更新を 5,000 回行った。

### 4.4 比較手法

本研究では EAP の効果を検証するために、以下の手法と比較した。

- (1) **Naive**：正則化手法を使用しないナイーブな AM
- (2) **AAP**：ユニットの出力値の代わりに、入力信号の振幅値を用いて計算する手法。AAP は「Abnormal Amplitude Penalty」の略である。AAP では閾値  $th^{(l,w)}$  をユニットごとに決めることができない。そのため、すべてのデータセットの分布を考慮し一律  $th^{(l,w)} = 10.0 \text{ G}$  とした。
- (3) **L2**：L2 正則化を用いた手法
- (4) **TV**：Total Variation を用いた手法
- (5) **EAP**：提案手法

式 (1) における正則化項の強さを決める重み  $w$  は、すべての正則化手法において  $w = 0.5$  とした。

### 4.5 評価指標

本研究は可視化研究であるため、まず定性的評価を行う。これに加えて、多くの可視化研究では行われていない定量的評価を行う。画像分野の可視化研究では、高周波成分ノイズの除去や本研究同様に実在のデータに近い信号を生成することが目的となるが、多くの研究では評価は定性的な評価にとどまる。生成した信号の評価としては、GAN における Inception Score [17] などがあるが、生成された信号の多様性などが評価軸としてあり、AM の評価に適用することはできない。したがって、本節では生成された波形と学習データを比較することで、生成されたデータが実在のデータに近いか否かを判断するための評価指標を提案する。

AM は、注目したユニットが抽出する学習データに存在する特徴的なパターンを入力信号として表現する手法であり、正則化項はノイズや異常値を抑制することで、AM の結果をより実在するパターンに近づけるために導入されている。したがって、本研究では AM の出力と学習データのサンプルを比較し、学習データに類似するセグメントを生成できているか調べることで、正則化手法の評価を行うことができると思う。AM の出力  $\mathbf{X}^*$  には、抽出された特徴的な波形が繰り返し出現するが、実際のデータにはこの特徴的な波形は繰り返し出現するとは限らない。そこで、まず AM の結果にウィンドウ幅の 10 pt (= 0.33 sec) のスライディングウィンドウを適用しセグメントに切り分け、その中からユニットの出力値が最も大きいセグメントを代表として選択する。次に、このセグメントとすべての学習サンプルの平均二乗誤差 (MSE) を計算し、MSE が小さい  $N$  件の学習サンプルを選択する。セグメントの長さは学習サンプルより小さいため、MSE は各学習サンプルの頭から 1 点ずつずらしながら計算する。最後に、上記の操作をすべてのユニットに対して実行し、それぞれのユニットに対して選択した  $N$  件の類似するサンプルとの MSE の平均を評価指標とする。本実験では  $N = 50$  とした。また、AM の出力  $\mathbf{X}^*$  の振幅が 10 G を超える場合は、類似サンプルを見つけることが難しい。このような結果は評価指標に大きな悪影響を与えるため、あらかじめ除外した。以後、この評価指標を MSE-AM と表記する。

### 4.6 結果

#### 4.6.1 定性的評価

図 2 に、データセット DSADS を学習したモデルに対して AM を適用した結果を示す。Naive, AAP, L2, TV の結果は比較的似通っており、ノイズを多く含み、最大振幅が 10 G を超えるような信号が生成されていることが分かる。青い背景で示されたユニットの出力値 (Activation)

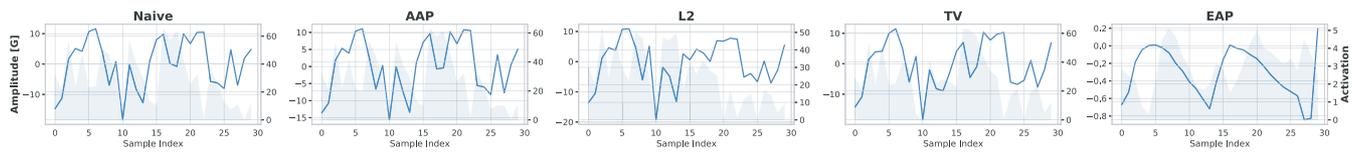


図 2 AM の正則化手法の比較 (定性的評価, DSADS 3 層目 Unit31) : 左の縦軸は波形の振幅を表し, グラフ中の青い背景と右の縦軸は対応するユニットの出力値  $X^{(l,u)}$  を表す

Fig. 2 Example of generated signals by different regularization methods. The horizontal axis shows the sample index. The background histogram and right vertical axes show the activation values of the corresponding samples.

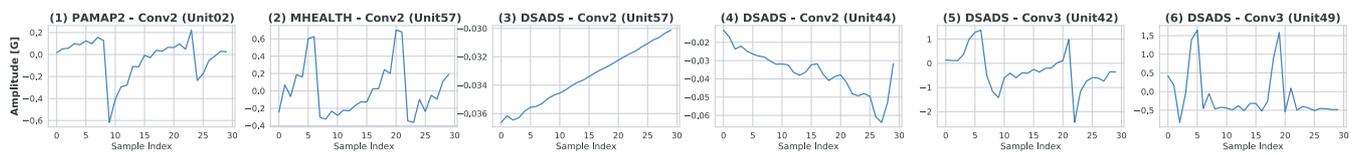


図 3 提案手法 (EAP) によって生成された信号

Fig. 3 Example of signals generated by the proposed method.

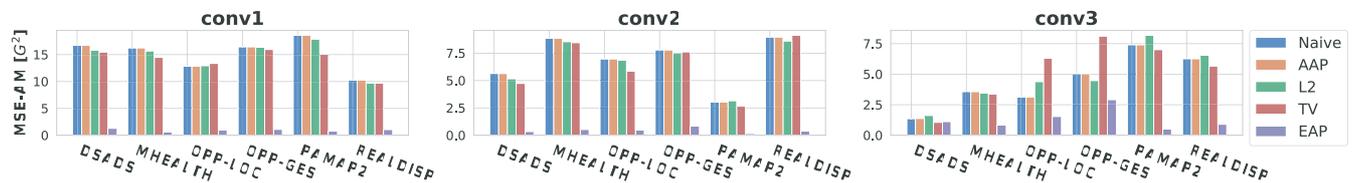


図 4 AM の正則化手法の比較 (MSE-AM による定量的評価)

Fig. 4 Result of quantitative analysis by MSE-AM.

も, 学習データの最大値は 14.8 であるが, 60 を超えるような非常に大きな値が出現している. これに対して, EAP を用いた手法はノイズが少なく, 右下がりの波形を抽出しているユニットだと分かる. ユニットの出力値も他の手法に比べて非常に小さく, 学習データの分布に近い結果となった.

L2, TV はユニットの特徴や生成している信号の状況に合わせて正則化の強さを変えることができない. 最適化の繰り返し処理の過程で, ユニットの出力値と振幅の値は必ずしも同時に大きくなるわけではない. そのため, 振幅の値の変化に敏感に反応するユニットでは, 振幅の変化によってペナルティが大きくなるよりも, ユニットの出力値が早く大きくなり, この 2 つの値の差が拡大し続けることで正則化項がほとんど無視される結果になったと考えられる. 一方で EAP は, ユニットの出力値から正則化項を計算し, ユニットの出力値が大きくなるタイミングに応じて, ペナルティをかけ続けることができたため, 他の手法より効果が大きかったと考えられる.

図 3 に EAP を用いて AM を実行した結果を示す. 2 層目のユニットでは上向きや下向きの円弧が抽出されていることが, それぞれ図 3(1), (2) の Index 10-20 の波形から分かる. また, 図 3(3), (4) では, 2 層目のユニットにおいて右上がり・右下がりの波形が抽出されていることが確認できる. 3 層目のユニットでは, 緩やかな右上がりの後に急激に値が小さくなる波形や, 高い山の後に小さな山が

続く波形が捕らえられていることが, 図 3(5), (6) から分かる.

#### 4.6.2 定量的評価

図 4 に, 6 種類のデータセットに AM を適用し, 4.5 節で示した方法で計算した MSE-AM に基づくスコアを示す. 層によって MSE-AM の値が異なるため, MSE-AM は層ごとに計算した. グラフから分かるように EAP は, 他の手法と比較して非常に大きく MSE-AM を削減しており, 学習データに出現するパターンにより近い信号を生成できたといえる.

全体として, 層が深くなるほど MSE-AM が小さくなる傾向がある. 層が深くなるほど信号の生成が難しくなり, 異常に大きな振幅を持つ信号が生成されやすくなる. 評価指標はこれらの信号を除外して計算するため, MSE-AM が小さくなっている. DSADS, OPP-GES, OPP-LOC では, EAP を用いないと半数を超えるユニットで信号の生成に失敗 (振幅が 10 G を超える) しているが, EAP ではすべてのデータセットにおいて生成に失敗した信号はなかった.

Naive と AAP はほとんど MSE-AM の値が同じである. 振幅が AAP の閾値を超えてペナルティが大きくなり始めたとしても, ユニットの出力値はそれ以上に早く大きくなり, AAP が無視されていたことが考えられる. そのため, AAP はほぼ最初から効果がなく, Naive と非常によく似た結果になったと考えられる.

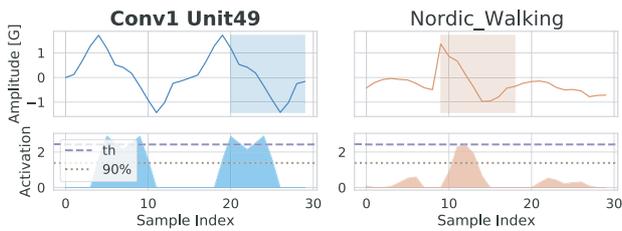


図 5 AM の結果に類似する学習データ：PAMAP2 1 層目ユニット 49 に対する AM の結果とそれに類似した学習サンプル。上段は波形を表し、背景の矩形は MSE-AM が最小となったセグメントである。下段は上段の波形に対応するユニットの出力値の分布である。th と 90% は、それぞれ EAP の閾値  $th^{(l,u)}$  と、すべての学習データを入力した際の注目したユニットの 0 以外の出力値の累積 90% の値である

Fig. 5 Signals generated by the proposed method (blue signals) and their most similar training segments (orange signals); the corresponding activation values are shown in the lower panels. The first dotted line “th” is the threshold value of the EAP. The second dotted line “90%” indicates the 90th percentile of the distribution of activation values for the corresponding unit when feeding all the training data.

#### 4.6.3 類似サンプルの探索

図 5 に、提案手法で生成した信号（左）と評価指標（MSE-AM）によって選択された類似サンプルの例（右）を示す。ただし、青色と橙色の背景で強調されている区間が類似サンプル探索の際に互いに最も類似していると判定された区間である。このユニットは図 5 の左下のユニットの出力値の分布から、右下がりの波形を検出するユニットだと考えられる。類似サンプルの右下がりの波形では、対応するユニットの出力値が大きくなっている。この結果から提案手法による可視化結果が、注目したユニットが抽出する学習サンプルに存在するパターンを反映していることが分かる。

またこの結果から、評価指標である MSE-AM が適切に類似サンプルを抽出できていることも確認できる。

### 5. 可視化結果に基づく転移元データセットの推薦

本章では提案した可視化手法を利用し、転移学習における転移元データセットを推薦する手法について説明し、4 種類のデータセットを用いて検証する。

本研究における転移学習では、以下の条件を想定する。

- 転移元・転移先のデータセットには、共にラベル付きのデータが存在する。ただし、転移先環境のデータは少ない。
- 転移元データセットの候補が複数あり、この中から認識精度が高くなる転移元データセットを選択したい。

本研究では転移元データセットとして、4.1 節に示した 6 種類を用いる。これらの転移元データセットを学習

したモデルには、4.2 節で学習したものを用いる。転移先データセットとして、DSADS, MHEALTH, PAMAP2, REALDISP を用いた。OPP-LOC, OPP-GES は被験者数が非常に少ないため、本節の検証からは除外した。転移先のデータセットに関して、学習データとテストデータの分割は次のように決定した。まずテストデータとしては、4.2 節と同様に全体に対する 2 割の被験者を用いた。本研究では転移先の学習データが少ないことを想定している。学習データとしては、表 1 に示した「学習に使用した被験者数」のうち約半数の被験者をランダムに抽出することで作成した。ただし、半数の被験者の抽出方法は複数あり、本研究ではそれぞれを個別の転移先環境と見なす。したがって被験者の抽出方法を重複なく変えることで、DSADS から 20 件、MHEALTH から 70 件、PAMAP2 から 35 件、REALDISP から 50 件のデータセットを用意し、合計 175 の転移先データセットに対して、転移元データセットの推薦と転移学習を行った。認識精度 (F 値) の計算には、各転移先・転移元の組合せに対して 5 回ずつ学習を行いその平均値を用いた。本研究ではこのようにして、データが少ない環境での転移学習の評価を行う。転移学習は、転移元モデルの畳み込み層の重みを固定し、全結合層のみを転移先の少数データを用いて学習係数  $10^{-4}$  で 100 エポックほど学習する。

転移元と転移先の行動クラスは重複がないものが多いが、すべてのデータセットに共通する行動クラスのみを抽出して利用するなどといった、利用する行動クラスの選択はない。これは、同じ名前の行動ラベルを持つ異なるデータセットのデータが必ずしも共通の特徴を持つとは限らないためである。本研究では行動クラスに依存しない転用可能な特徴が存在すると仮定し、AM を用いることでデータに含まれる特徴の観点から、適切な転移元データセットを選択することを目指す。

#### 5.1 転移元データセット推薦手法

転移先のデータで高い認識精度を達成するためには、転移先に存在する特徴を畳み込み層で抽出できることが重要だと考えられる。AM は、転移元となるモデルが畳み込み層で抽出する特徴表現を入力信号として表現する。したがって、転移元モデルの AM の結果と転移先データの類似度を 4.5 節で導入した MSE-AM を用いて計算し、類似度の高い転移元モデルを転移学習に用いればよい。

提案手法は以下の手順となる。(1) まず候補となる転移元モデルの最も深い畳み込み層に対して AM を実行する。(2) 次に転移元モデルの AM の結果と、転移先のデータに対して MSE-AM を計算する。(3) 最後に MSE-AM が小さい順に転移元モデルを推薦する。

5.2 評価方法

5.2.1 比較手法

Wang ら [19] は、本研究同様に転移元データセットの推薦を行う手法を提案した。この手法は転移元データと転移先データ間の距離を、ドメイン分類精度から計算される  $\mathcal{A}$ -distance [4] と、データセットのセマンティクスなどを考慮した距離の 2 つを用いて測る。2 つ目の指標は研究者の主観による重みを用いるため、本研究では客観的な 1 つ目の指標のみを用いる。

$\mathcal{A}$ -distance の計算方法は以下のとおりである。まず転移元データセットと転移先データセットからそれぞれ  $m'$  サンプルを抽出し、これを入力としてどちらのデータセットに属するかを推定する 2 値分類器を学習する。次に分類誤差  $\epsilon$  を計算し、以下の式に代入することで  $\mathcal{A}$ -distance ( $D_{\mathcal{A}}$ ) を得る。

$$D_{\mathcal{A}} = 2(1 - 2\epsilon) \tag{5}$$

$\mathcal{A}$ -distance は小さいほど、転移元と転移先のデータが類似していることを示すため、 $\mathcal{A}$ -distance が小さい順に転移元データセットを推薦する。

本研究では MSE-AM に基づく提案手法、 $\mathcal{A}$ -distance を用いた手法、ランダムな順番で推薦を行う手法の 3 つを比較する。

5.3 結果

転移元データセットを推薦手法は、候補となる転移元データセットに対して、転移学習後の認識精度が高くなると推定される順にランキングを作成する。そこで、まず作成されたランキングの評価を行う。評価指標として、平均二乗誤差 (MSE) を用いた。MSE は正解および推定ランキング間で計算し、値が小さい方が、転移学習後の認識精度が高い順にデータセットを推薦できていることを示す。結果を表 2 に示す。提案手法 (MSE-AM) の MSE は、比較手法に対して約 40%削減している。この結果から、提案手法は他の手法より高い精度で、転移学習後の認識精度の順番を推定できていることが分かる。

次に、転移元データセットを推薦することによって、F1 値がどの程度最良の結果に近くなるかを評価するため、各転移先データセットで最良の F 値に対する 1 位に推薦したデータセットの F 値の差分を計算した。この F 値の低下幅をデータセットごとに集計し、平均値と最悪値を表 3 に示す。数値は小さい方が最良の結果からの低下が小さいことを示す。低下幅の平均値に関して、MHEALTH では

表 2 転移元データセットの推薦ランキングの評価

評価指標	Proposed (MSE-AM)	$\mathcal{A}$ -distance	Random
MSE	2.48	4.151	4.000

$\mathcal{A}$ -distance に対して提案手法の方が低下幅が大きくなっているが、残りの 3 つでは大きな差はない。しかし、低下幅の最悪値は提案手法の方がすべてのデータセットに対して小さくなった。

図 6 に、提案手法と  $\mathcal{A}$ -distance の 1 位に推薦したデータセットの F 値、および候補の中で最も低いデータセットの F 値 (Worst Case) に対する最良の F 値からの低下幅の分布を示す。図 6(1) より、認識精度の低下幅は提案手法 (MSE-AM, 縦軸) ではほとんど 0.10 程度に抑えることができているが、 $\mathcal{A}$ -distance を用いた手法では 0.10 以上認識精度が低くなるデータセットを推薦している場合が多くみられる。 $\mathcal{A}$ -distance で 1 位に推薦したデータセットを用いると、最も高い F 値に対して 0.18 以上認識精度が低くなるデータセットを推薦している。これに対して提案手法は、悪くても 0.09 程度の F 値の低下に抑えている。また図 6(2) に示すように、データセットをランダムに使用すると、F 値が 0.20 以上小さくなるデータセットを選択する可能性があるが、提案手法を用いることで低下幅を大きく抑えていることが分かる。

以上の結果より、提案手法は比較手法より転移元データセットを推薦する際に、認識精度が低くなるような推薦を行うリスクを低減できていることが分かる。提案手法では、

表 3 推薦 1 位を用いたときの F 値の低下幅 (平均・最悪) : 低下幅が小さい結果を太字で示す

Table 3 Decrease in F-measure (average and worst) when using the highest ranked recommendation: Results with smaller decrease are shown in bold.

-	推薦手法	DSADS	MEHALTH	PAMAP2	REALDISP
平均	Proposed	0.028	0.074	<b>0.015</b>	<b>0.053</b>
	$\mathcal{A}$ -distance	<b>0.023</b>	<b>0.029</b>	0.027	0.077
最悪	Proposed	<b>0.081</b>	<b>0.125</b>	<b>0.057</b>	<b>0.103</b>
	$\mathcal{A}$ -distance	0.093	0.173	0.083	0.185

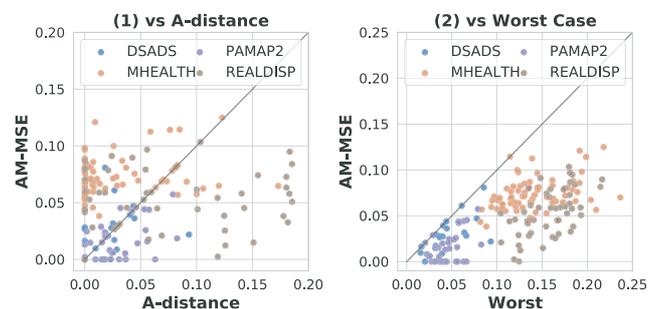


図 6 F 値の減少幅の比較: (1) 提案手法と  $\mathcal{A}$ -distance の 1 位の比較, (2) 提案手法 1 位と候補の中で最も低い F 値の比較。ただし各点はそれぞれの転移先環境に対応する

Fig. 6 Comparison of decrease in F-measures: (1) comparison between the first place of the proposed method and that of  $\mathcal{A}$ -distance, (2) comparison between the first place of the proposed method and the lowest F-measure among the candidates.

実際にモデルが学習した特徴を用いることで、行動の認識には直接寄与しないデータセットの特徴が、データセットの類似度計算に与える影響を抑えることができる。そのため、提案手法はデータの特性しか考慮しない比較手法を上回ったと考えられる。

## 6. 考察

4.3 節で説明したように、評価実験では1軸のセンサデータを対象にAMを実行した。しかし提案手法は、複数のセンサに対しても適用可能である。たとえば、畳み込み層において入力全体やセンサごと(XYZの3軸)から特徴抽出するモデルに対して、AMを適用することも可能である。この場合、最適化する信号はそれぞれ $(N_S \times N_T) = (6 \times 30)$ ,  $(3 \times 30)$ の2次元のテンソルとなる。複数センサから特徴抽出を行う場合、注目したユニットが特に重視しているセンサなどを分析することができる。ただし、最適化対象の値が増加するため最適化が難しくなり、学習係数などのパラメータを適切に設定することが重要となる。

## 7. 結論

本研究では行動認識分野における転移学習の活用を推進すべく、行動認識モデルの分析のための可視化手法と、可視化結果を用いて転移元データセットの推薦手法を提案した。行動認識モデルが抽出している特徴量を可視化するために、画像処理分野で用いられる Activation Maximization (AM) と呼ばれる手法を行動認識モデルに適用し、注目したユニットが最も強く応答する入力信号を生成することを試みた。行動認識モデルの入力となる加速度信号は画像に比べて値域が広く、信号を生成する過程で異常値が発生する。この問題に対し、ユニットの出力値を用いて正則化を行う Extreme Activation Penalty (EAP) を提案した。また、可視化手法の応用例として、転移学習の転移元モデルの推薦手法を提案した。提案手法では、AMの結果と転移先データを比較することで、転移元モデルが抽出する特徴と転移先データに存在する特徴が最も近いデータセットを推薦する。本研究では、可視化手法と転移元モデルの推薦手法の有効性を、公開データセットを用いて確認した。今後は、推薦手法の改善を行うとともに、より多くのデータセットを用いて検証をしていきたい。

**謝辞** 本研究の一部は JST CREST JPMJCR15E2, JSPS 科研費 JP16H06539, JP17H04679 の助成を受けて行われたものである。

## 参考文献

- [1] Baños, O., Damas, M., Pomares, H., Rojas, I., Tóth, M.A. and Amft, O.: A benchmark dataset to evaluate sensor displacement in activity recognition, *Proc. 2012 ACM Conference on Ubiquitous Computing*, pp.1026–1035 (2012).
- [2] Banos, O., Garcia, R., Holgado-Terriza, J.A., Damas, M., Pomares, H., Rojas, I., Saez, A. and Villalonga, C.: mHealthDroid: A novel framework for agile development of mobile health applications, *Proc. International Workshop on Ambient Assisted Living*, pp.91–98 (2014).
- [3] Barshan, B. and Yükek, M.C.: Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units, *The Computer Journal*, Vol.57, No.11, pp.1649–1667 (2014).
- [4] Ben-David, S., Blitzer, J., Crammer, K. and Pereira, F.: Analysis of representations for domain adaptation, *Proc. Advances in Neural Information Processing Systems*, pp.137–144 (2007).
- [5] Erhan, D., Bengio, Y., Courville, A. and Vincent, P.: Visualizing higher-layer features of a deep network, *University of Montreal*, Vol.1341, No.3, p.1 (2009).
- [6] Khan, M.A.A.H., Roy, N. and Misra, A.: Scaling human activity recognition via deep learning-based domain adaptation, *Proc. 2018 IEEE International Conference on Pervasive Computing and Communications*, pp.1–9 (2018).
- [7] Le, Q.V., Ranzato, M., Monga, R., Devin, M., Corrado, G., Chen, K., Dean, J. and Ng, A.Y.: Building high-level features using large scale unsupervised learning, *Proc. 29th International Conference on Machine Learning* (2012).
- [8] Maas, A.L., Hannun, A.Y. and Ng, A.Y.: Rectifier nonlinearities improve neural network acoustic models, *Proc. International Conference on Machine Learning*, Vol.30, No.1, p.3 (2013).
- [9] Mahendran, A. and Vedaldi, A.: Understanding deep image representations by inverting them, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.5188–5196 (2015).
- [10] Münzner, S., Schmidt, P., Reiss, A., Hanselmann, M., Stiefelhagen, R. and Dürichen, R.: CNN-based sensor fusion techniques for multimodal human activity recognition, *Proc. 2017 ACM International Symposium on Wearable Computers*, pp.158–165 (2017).
- [11] Nguyen, A., Clune, J., Bengio, Y., Dosovitskiy, A. and Yosinski, J.: Plug & play generative networks: Conditional iterative generation of images in latent space, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.4467–4477 (2017).
- [12] Nguyen, A., Dosovitskiy, A., Yosinski, J., Brox, T. and Clune, J.: Synthesizing the preferred inputs for neurons in neural networks via deep generator networks, *Proc. Advances in Neural Information Processing Systems*, pp.3387–3395 (2016).
- [13] Nguyen, A., Yosinski, J. and Clune, J.: Deep neural networks are easily fooled: High confidence predictions for unrecognizable images, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.427–436 (2015).
- [14] Reiss, A. and Stricker, D.: Introducing a new benchmarked dataset for activity monitoring, *Proc. 2012 16th International Symposium on Wearable Computers*, pp.108–109 (2012).
- [15] Roggen, D., Calatroni, A., Rossi, M., Holleczeck, T., Förster, K., Tröster, G., Lukowicz, P., Bannach, D., Pirkel, G., Ferscha, A., et al.: Collecting complex activity datasets in highly rich networked sensor environments, *Proc. 2010 7th International Conference on Networked Sensing Systems*, pp.233–240 (2010).
- [16] Saks, S.: Theory of the integral, 2nd ed., English trans-

lation by Young, L.C., with two additional notes by Banach, S., Vol.7 (1937).

- [17] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A. and Chen, X.: Improved techniques for training gans, *Proc. Advances in Neural Information Processing Systems*, pp.2234–2242 (2016).
- [18] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D.: Grad-CAM: Visual explanations from deep networks via gradient-based localization, *Proc. IEEE International Conference on Computer Vision*, pp.618–626 (2017).
- [19] Wang, J., Zheng, V.W., Chen, Y. and Huang, M.: Deep transfer learning for cross-domain activity recognition, *Proc. 3rd International Conference on Crowd Science and Engineering*, p.16 (2018).
- [20] Yosinski, J., Clune, J., Nguyen, A.M., Fuchs, T.J. and Lipson, H.: Understanding neural networks through deep visualization, *CoRR*, Vol.abs/1506.06579 (2015).

### 推薦文

本論文はユビキタスコンピューティングと機械学習の融合研究であり、ユビキタスコンピューティングにおいて学習データを収集することの難しさと、機械学習において要因を分析することの難しさを、データの可視化と転移学習への応用によって解決する斬新なアイデアを提案し、入念に評価をしている。

(ユビキタスコンピューティングシステム研究会主査  
井上 創造)



原 隆浩 (正会員)

1995年大阪大学工学部情報システム工学科卒業。1997年同大学大学院工学研究科博士前期課程修了。同年同大学院工学研究科博士後期課程中退後、同大学院工学研究科助手、2002年同大学院情報科学研究科助手、2004年同大学院情報科学研究科准教授。2015年より同大学院情報科学研究科教授となり、現在に至る。工学博士。2003年本学会研究開発奨励賞受賞。2008年、2009年本学会論文賞、2015年日本学術振興会賞受賞。2017年大阪科学賞受賞。ソーシャルコンピューティング、ネットワーク環境におけるデータ管理技術に関する研究に従事。IEEE, ACM, 電子情報通信学会各会員。



吉村 直也 (正会員)

2018年大阪大学工学部電子情報工学科卒業。2020年同大学大学院情報科学研究科博士前期課程修了。ユビキタスコンピューティング、ウェアラブルセンシングに関する研究に従事。



前川 卓也 (正会員)

2006年大阪大学大学院情報科学研究科博士後期課程修了。2006年NTTコミュニケーション科学基礎研究所入所。2012年より大阪大学大学院情報科学研究科マルチメディア工学専攻准教授、現在に至る。ユビキタスコンピューティング、ウェアラブルセンシングに関する研究に従事。2010年度情報処理学会山下記念研究賞、2013年度日本データベース学会上林奨励賞、2015年電気学会優秀論文発表A賞、2019年IPSJ/IEEE Computer Society Young Computer Researcher Award等受賞。IEEE, ACM, 電気学会, 日本データベース学会各会員。