半教師あり蒸留による深層学習に基づく行動認識モデルのユーザ適応

Semi-Supervised Distillation: Personalizing Deep Neural Networks in Activity Recognition using Inertial Sensors

岩澤 有祐
Yusuke Iwasawa

東京大学大学院工学系研究科技術経営戦略学専攻
Graduate School of Engineering, The University of Tokyo
iwasawa@weblab.t.u-tokyo.ac.jp

矢入 郁子
Ikuko Eguchi Yairi

上智大学理工学部情報理工学科
Graduate School of Science and Technology, Sophia University
i.e.yairi@sophia.ac.jp

松尾 豊
Yutaka Matsuo

東京大学大学院工学系研究科技術経営戦略学専攻
Graduate School of Engineering, The University of Tokyo
matsuo@weblab.t.u-tokyo.ac.jp

keywords: personalization, deep neural networks, action recognition, distillation, semi-supervised

Summary

Personalization of activity recognition has become a topic of interest to improve recognition performance for diverse users. Recent researches show that deep neural networks improve generalization performance in activity recognition using inertial sensors, such as accelerometers and gyroscopes; however, personalizing deep neural networks is challenging because it has a thousands or millions of parameters but generally personalization should be done with small amount of labeled data.

This paper proposes novel way to personalize deep neural networks by preventing overfitting using un-labeled data. This is done by adding output-distribution similarity regularization between the reference model and personalized models, which is an extension of distillation recently proposed by Hinton. Experiments on an opportunity activity recognition dataset, one of the most famous datasets in the fields, demonstrates that the proposed regularization techniques prevent overfitting even if we have few labeled data for each target classes per users, and provide better recognition performances compared with other personalization techniques. We also conduct various experiments, including a no-labeled data setting and combination of the proposed method and well-used personalization techniques to check if the proposed method is complementary with existing methods or competitor of its. The results suggest that the proposed regularization works well in various settings, and complementary with existing methods.

1. はじめに

ウェアラブルセンサによる人間行動の認識と計測（以降行動認識）は、高齢者の転倒検出などの見守りシステム、健康管理など種々のアプリケーション実現のための中心的な技術要素である。画像認識や音声認識などの領域と同様、行動認識の分野でも Deep Neural Networks (DNN) が高い認識精度を示すことが近年の研究で示されている [Hammerla 16, Plötz 11, Yang 15]。行動認識でよく利用される加速度センサやジャイロスコプなどの慣性センサは生データに近く人間による特徴設計が難しいこと、行動と一口に言えども認識したい対象はアプリケーションによって異なること、それゆえアプリケーションごとに有効な特徴量は異なることなどから、DNN は今後も重要度を増していくものと考えられる。

DNN は行動認識の精度を高めるための重要な技術要

素となっている一方で、行動認識に利用する場合にはいくつかの未解決問題がある。その1つが、ラベルデータが限られたユーザに対する認識精度の問題である。行動認識でよく利用される慣性センサへの観測値のみによるアノテーションは困難であり、主に動画データを利用したアノテーションが行われる。しかし、プライバシーやユーザ負担の観点から研究室などの実験環境以外でラベルデータを大量に取得するのは困難である。そのため、ラベルデータが少ない、あるいはまったくないユーザの行動を認識する方法が必要となる。

本稿ではこの問題の解決に向けて、DNN を利用して得られた行動認識モデル（以降、深層行動認識モデル）をユーザ適応する方法を提案する。ここで、行動認識モデルのユーザ適応とは、不特定多数のユーザのサブセット $\{(X_i, y_i)\} (\forall i \in 1 \dots m)$ で訓練された不特定行動者モデルをユーザごとに用意した少数の適応用データ $\{(X', y')\}$

図 1: ユーザ適応の概念図。

を利用して、特定のユーザの行動をよく認識するようなモデル（以降、適応後モデル）へと修正することを指す（図 1）。言い換えれば、ユーザ適応は別のユーザ群のデータで学習された知識を利用することでデータが限られたユーザに対する認識精度を高めるための方法であり、決定木など単純な認識モデルにおける有用性が報告されている [Zhao 11]。一方、決定木のようにパラメタの少ないモデルをユーザ適応する場合と異なり、DNN のように大量のパラメタを持つモデルをユーザ適応する場合どう過学習を防ぐかが重要な論点となる。

本稿では、蒸留 [Hinton 15] を拡張した半教師あり蒸留によるユーザ適応手法を提案する。蒸留とは、学習済みモデル（教師モデル）の出力を温度パラメタ T により滑らかにしたものを模倣するように新たなモデル（生徒モデル）を学習させる方法である。提案手法の背景にある基本的なアイデアの 1 つは、ユーザ適応後のモデルは不特定行動者モデルから過剰に離れるべきではないというものである。もし過剰に離れることを許容するならば、不特定行動者モデルを介した意味がないことになるためこれは自然な発想であろう。2 つ目のアイデアは、ラベルなしデータの活用である。行動認識では連続的にデータを計測する場合が多く、そのためラベルなしデータについては大量に得られる場合が多い。提案手法は不特定行動者モデルを教師モデルとし、適応後モデルを生徒モデルと考え、教師モデルを模倣する正則化項の計算にラベルなしデータを活用するように蒸留をユーザ適応の文脈に拡張する。提案手法は蒸留を上記のように拡張することで、1. 適応後モデルが不特定行動者モデルから離れすぎないように制約をかけられる、2. ラベルなしデータを自然に活用できる、という利点がある。

不特定行動者モデルから過剰に離れるのを防ぐという観点で、提案手法は音声認識の領域で提案されている KL 正則化 [Yu 13] と同様の目的を持つ。実際、蒸留による正則化は KL 正則化をその一部に含んでおり、パラメタ T を考慮することで一般化したものとみなせる。また、提案手法はラベルなしデータを活用するための仮定の 1 つである Low-Density Separation をユーザ適応の文脈で利用しているものとみなすことができる。このことから、ラベルなしデータを有効に活用できると期待される。

提案手法の有効性検証にあたり、深層学習を利用した行動認識モデルの評価で最もよく利用される Opportunity

データセットを利用して比較実験を行った。比較対象としては、音声認識の分野で提案されている Singular Value Decomposition (SVD) をベースにした手法 [Xue 14] および KL 正則化 [Yu 13] を含む計 8 種類のベースラインと比較し、提案手法により過学習が抑制されること、既存手法と比較して認識精度を改善できること、ラベルなしデータを有効に利用できていることを検証した。

本論文の主要な貢献は次のとおりである。

- DNN ベースの行動認識モデルを利用する場合のユーザ適応の有効性を初めて検証した。この研究は今後の深層行動認識モデルのユーザ適応技術に関する研究の第 1 歩となるものである。
- DNN ベースの行動認識モデルを少数のラベルデータでユーザ適応するための手法として、蒸留を応用した半教師あり蒸留 (Semi-Supervised Distillation: SSD) を提案した。本手法により、DNN のように大量のパラメタを持つモデルを利用する場合でもラベルデータが限定されたユーザに対する認識精度を効率的に高めることができる。
- 代表的なデータセットの 1 つである Opportunity データセットを利用して、複数の既存手法との比較実験を行い提案手法の有効性を定量的に検証した。

本論文の残りの構成は次の通りである。2 章では、まず行動認識での DNN の研究動向について概観し、DNN のユーザ適応に関する既存研究について言及する。また、提案手法の基礎となる蒸留についてその研究動向をまとめ、提案手法の新規性について主張する。3 章では、提案手法について説明する。4 章では、提案手法の有効性検証のために行った実験の詳細および実験結果を説明し、5 章で本論文で取り組んだ深層行動認識モデルのユーザ適応という問題に対する本稿の貢献および将来課題について述べる。最後に 6 章で本論文をまとめる。

2. 関連研究

2.1 行動認識での DNN の利用

近年、さまざまな研究において DNN の行動認識における有効性が検証され、DNN は高精度な行動認識のための必須の技術要素となりつつある。行動認識において DNN が重要な技術要素である理由の 1 つが特徴設計の難しさである。行動認識では加速度センサやジャイロセンサがよく利用されるが、これらのセンサは生の系列データであり人間がその特徴を識別するルールを記述するのは困難である。DNN などの表現学習技術によりデータから直接認識に寄与する特徴量を獲得することができれば、行動認識で経験的に利用されてきた特徴量では捉えきれなかった情報を識別に利用することができる。特徴量の設計は機械学習を適用する際に認識精度に影響を与える要素の 1 つであり、特徴設計の難しい行動認識における DNN の重要性の理由の 1 つとなっている。

Plötz らは行動認識に DNN をはじめて利用し、従来の特徴設計パイプラインを RBM や PCA で置き換える方法の有効性を 4 つのデータセットで検証した [Plötz 11]。また、Vollmer らはスパース符号化を利用したニューラルネットワークの有効性を報告した [Vollmer 13]。近年の研究で、Jing らは畳み込みニューラルネットワークが行動認識タスクと親和性が高く、認識精度も高いことを報告している [Yang 15]。また、Ordóñez や Hammerla らが Long Short Term Memory のように時系列性を考慮した手法が良い精度を発揮することを確認している [Hammerla 16, Ordóñez 16]。これらのように、DNN の行動認識タスクにおける利用は着実に進んでおり、認識精度も改善し続けている。

2.2 深層行動認識モデルのユーザ適応

行動認識タスクにおける DNN の利用は活発に行われているものの、DNN で学習したモデルに対するユーザ適応の有効性の検討は行われていない。ここでは、DNN 以外のモデルにおける行動認識モデルのユーザ適応および関連領域である音声認識における DNN のユーザ適応に関する研究に言及する。

行動認識においてユーザ適応が必要な理由は、どのユーザに対しても利用可能な汎用モデルを作成することの難しさにある。例えば Lane らは、50 名分の行動データセットを利用して 20 代、30 代の人々が歩いている時と 65 歳以上の人々が歩いている時の特徴量の値がほとんど重ならないこと、また 20 代、30 代のデータで訓練したモデルは 20 代、30 代のユーザに対しては 80% から 90% の精度を達成する一方で 65 歳以上に対しては 60% 程度であったことを報告し、単一のモデルは全ユーザに対して適合させることは難しいことを指摘している [Lane 11]。また Weiss らは、特定のユーザのデータのみを利用して作成したモデルと、特定のユーザのデータを利用せずに作成したモデルの 2 つの認識精度を比較し、後者の認識精度が約 20% 低下することを報告している [Weiss 12]。これらの研究は、単純に学習したモデルは学習に利用したユーザに対しては良く認識できるものの学習時に存在しなかったユーザに対してはうまく認識できないことを示唆している。

ユーザ適応はこの問題に対処する方法の 1 つとして近年注目を集めている手法である。Zhao らはあるユーザのデータで学習された決定木の分類規則を学習時に存在しなかったユーザのデータで修正することで認識精度の改善が見られることを報告している [Zhao 11]。Ioffe らはパラメータをラベルなしデータのみから修正する方法として、共変量シフト仮定のもとサンプルの重要度を計算して学習させる方法を提案している [Ioffe 15]。Reiss らは、アンサンブル学習した複数の分類器の重みを修正することで効率的にユーザ適応を行う方法を提案している [Reiss 13]。一方、これらの手法は決定木に特化した更新

方法となっている、新規ユーザが現れるたびにモデル全体を学習し直す必要がある、大量のパラメータを持つ DNN をアンサンブル学習させるのは難しいなどの理由から DNN をユーザ適応する上では適した方法ではない。

ユーザ適応は従来、行動認識よりむしろ音声認識の分野で研究が進められてきた。音声認識におけるユーザ適応の代表的な手法の 1 つとして i-Vector [Dehak 11] などユーザを表すベクトルを補助情報として入力する手法が提案されている [Karanasou 14, Saon 13]。しかし、i-Vector のように音声認識においてユーザを表すベクトル表現として確立された手法は行動認識においては存在しておらず、直接同様の方法を利用することはできない。

音声認識におけるユーザ適応のその他の主要なアプローチの 1 つに、限定したパラメータのみをユーザ適応するモデルベースアプローチがある。例えば、Xue らは学習させるパラメータを減らす方法の 1 つとして重み行列を SVD により行列分解し、パラメータ数を減らした上でユーザ適応させる方法を提案している [Xue 14]。また、Yao らは出力層のバイアスを補正する方法を提案している [Yao 12]。Siniscalchi らは、活性化関数にパラメータを用意しユーザごとに活性化関数の形を変化させる方法を提案している [Siniscalchi 13]。これらのように適応させるパラメータを減らすモデルベースアプローチ自体はさまざまな研究がなされており、これらの手法は音声認識に特化したものではなく行動認識の問題においても適用可能である。しかし、元の不特定行動者モデルから極端に離れることを防ぐ提案手法とは直交する概念であり、組み合わせて利用することが可能である。

2.3 蒸留によるユーザ適応

提案手法は、蒸留 [Hinton 15] をユーザ適応の文脈で利用するために拡張したものである。蒸留とは、学習済みのモデル（教師モデル）の確率分布を滑らかにしたものを模倣するように正則化を加えた新たなモデル（生徒モデル）を学習させる方法である。ここで、確率分布を滑らかにするとは、例えば確率分布が $[0.99, 0.008, 0.002]$ といった確率になっていた場合に、 $[0.70, 0.20, 0.10]$ というようにする処理のことをさす。すなわち、蒸留では生徒モデルを学習する際に、ラベルデータから求められる損失に加えて元の教師モデルの出力を模倣するような制約をかけて学習する。このように、2 段階の学習を行うことで、より小さなニューラルネットワークで同等の認識精度を得られたり [Hinton 15]、より望ましい構造を持つニューラルネットワークに変換することができたり [Romero 15]、あるいは特定の条件下で半教師あり学習に有効である [Lopez-Paz 16] ことが報告されている。

ここでユーザ適応の観点から重要な点は次の 2 点である。第 1 に、教師モデルを不特定行動者モデルとし、生徒モデルを適応後モデルと考えることで、ユーザ適応に自然に利用できる点である。この場合蒸留による正則化は適応

後モデルが不特定行動者モデルから極端に離れないように制約を与えていると解釈することができる。第2に、単純な拡張によりラベルなしデータを活用できる点である。蒸留による正則化項は教師データの有無にかかわらず計算することができるため、ラベルなしデータを活用した学習ができる。詳細は3章にて述べるが、第2項の正則化は半教師あり学習でラベルなしデータを活用する際の主要な仮定の1つである Low-Density Separation[Chapelle 05]を実現していると解釈することができ、効率的なラベルなしデータの活用方法であると期待される。

筆者の知る限り、本稿は蒸留をユーザ適応の文脈に応用する初めての研究である。最も近い研究として、Yuらは不特定コンテキストで訓練されたモデルの出力の確率分布と適応後モデルの出力分布の KL 正則化により制約をかけて特定のコンテキストへと適応させることを提案している[Yu 13]。また、上述した SVD による重み W を低ランク近似する研究[Xue 14]においても同様の技術が利用されている。KL 正則化はユーザ適応時に過学習するのを避けるために元のモデルから過剰にパラメタが更新されるのを避けるという点で提案手法と同一の目的を持っている。また提案手法のパラメタ $T = 1$ とし、損失関数として交差エントロピーを利用した場合に蒸留による正則化は KL 正則化と一致する。一方、KL 正則化の場合不特定行動者モデルの出力を完全に模倣しようとするため、制約が強すぎる可能性がある。特に、KL 正則化の場合不特定行動者モデルが誤っているサンプルについての影響が強く残ることになり望ましくない場合がある。蒸留を利用した方法では温度 T を変化させることで不特定行動者モデルにおいて確信度が低いサンプルについてはその重みを低下させることができるため、KL 正則化より良い制約となると考えられる。

3. 半教師あり蒸留によるユーザ適応

3.1 アルゴリズム

提案手法である半教師あり蒸留 (SSD) によるユーザ適応は、蒸留をユーザ適応の文脈に合うように、またラベルなしデータを活用できるように応用したものである。提案手法は不特定行動者モデルの学習および SSD を利用したユーザ適応の2つのステップから構成される。順に説明する。

ステップ 1: 不特定行動者モデルの学習

ステップ 1 では、教師付きの全データ $\{X^j, y^j\} (\forall j \in 1 \dots m)$ を利用して不特定行動者モデルを学習する。ここで、 X^j および y^j はそれぞれユーザ j の入力データとラベルデータ、 m は全ユーザ数である。この最適化問題の損失は次のように定式化することができる。

$$L_\theta = \sum_{j=1}^m \frac{1}{n^j} \sum_{i=1}^{n^j} L(y_i^j, \sigma(f_\theta^t(x_i^j))) + \Omega(\|f_\theta\|), \quad (1)$$

θ は任意の構造を持つニューラルネットワークのパラメタであり、式 1 を確率的最急降下法などにより最小化することで求める。ここで、 x_i^j はユーザ j の i 番目のサンプル、 f_θ は θ をパラメタに持つニューラルネットワーク、 n^j はユーザ j のサンプル数、 Ω は関数全体の形を制限する正則化項である。 σ はソフトマックス関数であり、ニューラルネットワークの出力を確率値へと変換している。また、 L は入力と出力の間の損失を定義する任意の関数であり、一般的には交差エントロピーが利用される。ここで重要な点は、不特定行動者モデルが訓練時に存在するユーザのみではなく未知のユーザに対してもある程度の認識精度を持つように学習することである。そのため、式 1 に示したように正則化を行うことで学習データ(ユーザ)に対する過学習を防いだり、アプリケーションを利用することが想定されるユーザを可能な限り包含したようなユーザのデータを学習に利用したりすることで特定のユーザにのみ識別能力が高くなりすぎないようにすることが望ましい。

ステップ 2: SSD を利用したユーザ適応

ステップ 2 では、新規ユーザのデータ $\{X', y'\}$ を利用して新規ユーザに対する認識精度を高めるようにパラメタ θ を修正する。ここで、新規データのうちラベル付きのサブセットを $\{X'_{sup}, y'_{sup}\}$ 、ラベルなしのデータセットを $\{X'_{unsup}\}$ とすると、本ステップで解く最小化問題の損失は次のようになる。

$$\begin{aligned} L_{\theta'} &= \sum_{x, y \in \{X'_{sup}, y'_{sup}\}} (1 - \lambda) L(y, \sigma(f_{\theta'}(x))) \\ &\quad + \sum_{x \in \{X'_{sup}\}} \lambda L(s, \sigma(f_{\theta'}(x))) \\ &\quad + \sum_{x' \in \{X'_{unsup}\}} \lambda L(s, \sigma(f_{\theta'}(x'))) \\ &= \sum_{x, y \in \{X'_{sup}, y'_{sup}\}} (1 - \lambda) L(y, \sigma(f_{\theta'}(x))) \\ &\quad + \sum_{x \in \{X'\}} \lambda L(s, \sigma(f_{\theta'}(x))) \\ &= (1 - \lambda) \sum_{x, y \in \{X'_{sup}, y'_{sup}\}} L_{hard} + \lambda \sum_{x \in \{X'\}} L_{soft} \end{aligned} \quad (2)$$

ここで、 θ' は任意の構造を持つニューラルネットワークのパラメタであり、 λ は第1項と第2項の重み付けをするハイパーパラメタである。なお、蒸留の元論文[Hinton 15]では $|\theta'| < |\theta|$ とした場合の認識精度の変化を確認するため θ' は θ とは異なる値に初期化されていたが、ここでは不特定行動者モデルをユーザに適した形で修正することが目的であるため θ' の初期値として θ をそのまま利用した。また、 s は $s = \sigma(\frac{f_\theta(x)}{T})$ 、により計算される値であり、入力サンプル x を教師モデル f_θ に入力した際に得られる確率分布 $P_\theta(y|x)$ を温度 T により滑らかにしたもの

である．このように，SSD ではラベルなしデータを含むユーザ適応データセット X' から計算される L_{soft} により不特定行動者モデルから過剰に離れるのを抑制する．

3.2 既存手法との関係

前述したように $T = 1$ かつ L として交差エントロピーを用いたときに L_{soft} は [Xue 14, Yu 13] で利用されている KL 正則化と一致する．実際に式を書き下すと次のようになる．

$$\begin{aligned} L_{soft} &= L(s, \sigma(f_{\theta'}(x))) \\ &= L(\sigma(f_{\theta}(x)), \sigma(f_{\theta'}(x))) \\ &= L(P_{\theta}(y|x), P_{\theta'}(y|x)) \\ &= \sum_{s=1}^S P_{\theta}(y = s|x) \log P_{\theta'}(y = s|x) \end{aligned}$$

ただし，ここで $P_{\theta}(y|x) = \sigma(f_{\theta}(x))$ であることを利用した．また， $T \rightarrow 0$ とした場合に s は Lee らが提案している擬似ラベルによる半教師あり学習 [Lee 13] で用いられるような確率 $P_{\theta}(y|x)$ が最も高い場所で 1 となり残りが 0 になるような One-Hot ベクトルになる．逆に $T = \infty$ とした場合， L_{soft} は定数とみなせるため単純な教師あり学習となる．このように，蒸留による正則化はソフトマックス関数における温度 T を考慮することで KL 正則化をはじめとするさまざまな手法を一般化したものとみなすことができる．

半教師あり学習の観点から考えると， L_{soft} は，半教師あり学習においてラベルなしデータを活用する主要な方法の 1 つである Low-Density Separation [Chapelle 05] と関係がある．Low-Density Separation とは，汎化能力を高めるような決定境界面は確率密度の低い領域に存在するという仮定であり，半教師あり学習のさまざまな手法で利用されている [Grandvalet 04, Rifai 11, Weston 12]． L_{soft} の項目はソフトターゲット s の空間における入力 x のエントロピーと対応しており， L_{soft} を最小化することは，ソフトターゲットの空間においてデータ点の密度が高い場所に決定境界面が引かれないようにすることと対応している．このように，提案手法は半教師あり学習においてラベルなしデータをうまく活用するための主要な仮定の 1 つと対応しており，ラベルなしデータをうまく活用することでより汎化能力の高い決定境界面を学習することができると期待できる．

これらの議論をまとめるに，提案手法は既存研究と次のような関係にある．

- (1) 提案手法は KL 損失や擬似ラベルなどの概念を温度 T を考慮することで一般化したものである．
- (2) 提案手法は半教師あり学習の考え方の 1 つである Low-Density Separation をユーザ適応の文脈で取り入れた手法である．

4 章の実験では，既存手法との対比に加えて，温度 T を変化させること，またラベルなしデータを活用することの有効性を検証する．

4. 実験

4.1 実験設定

§1 データセット

データセットとしては，Opportunity Recognition Dataset^{*1} (以下 Opportunity データセット) [Roggen 10, Sagha 11] を利用した．Opportunity データセットとは行動認識において最もよく利用されるデータセットの 1 つであり，朝食を作るシナリオにおける 18 種類^{*2}の行動データが 4 人分提供されているデータセットである．先行研究に従い [Yang 15]，ウィンドウサイズを 30，重なり率を 0.5 としたスライディングウィンドウ方式によりサンプルを生成し，4 人合計で 49,881 のサンプルを得た．

§2 ニューラルネットワークの構造

ニューラルネットワークの構造としては，不特定行動者モデルおよび適応後モデルの双方で畳み込みニューラルネットワークを利用した．なお，層の数やフィルタ数などのハイパーパラメタは，同じ Opportunity データセットを利用した研究である [Yang 15] を参考とした．本ネットワークは，1) 入力層，2) 3 回の畳み込み，プーリング，ReLU による活性化の繰り返し，3) 800 ユニットの全結合層，4) 出力層の合計 6 層のネットワークである．なお，畳み込みサイズ (フィルタ数) は順に $1 \times 5(50)$ ， $1 \times 5(40)$ ， $1 \times 3(20)$ とし，プーリングは最大プーリングを 1×2 ， 1×2 ， 1×1 で行った (図 A.1)．

§3 ベースライン

次の 8 種類のベースラインを提案手法と比較した．

- **Ref:** ユーザ適応を行う前の不特定行動者モデル．Ref より認識精度が高ければユーザ適応により認識精度が改善することを意味する．
- **Mixin:** ユーザ適応を行うのではなく，新たなユーザの教師データが得られた際にそのユーザのデータも加えて 1 からニューラルネットワークを学習させた．
- **Fine Tuning (FT, FT-T, FT-D):** 不特定行動者モデルを式 1 で学習した後に，蒸留による正則化を行わずに新規ユーザのデータで再学習させた．これは，式 2 の $\lambda = 0$ とした場合に対応する．また，適応後モデル学習の際にすべての層を再学習する場合 (FT)，入力側に近い畳み込み層のパラメタを固定とし出力側の全結合層 2 層のみを再学習する場合 (FT-T)，最終出力層のパラメタを固定とし畳み込み層および全結合層を再学習する場合 (FT-D) の 3 種類を検討した．

^{*1} <http://opportunity-project.eu/challenge>

^{*2} open door 1, open door 2, close door 1, close door 2, open fridge, close fridge, open dishwasher, close dishwasher, open drawer 1, close drawer 1, open drawer 2, close drawer 2, open drawer 3, close drawer 3, clean table, drink cup, toggle switch, and Null

- **SVD**[Xue 14]: Xue らにより提案された手法をベースラインとして利用した．全結合層および出力層の重み行列 W を SVD により $W = USV$ と分解し，特異値行列 S を再学習した．ここで，上位 k 個の特異値を持つ要素のみを残して近似を行った場合に， S の持つ要素数は $2k$ となり，もとの行列 W の要素数 mn より小さくなることに注意されたい．なお，他に言及がない限り本稿では $k = \min(m, n)$ とした．
- **Supervised Distillation (SD)**: ラベルなしデータは利用せず，ラベルありデータのみを利用して蒸留による正則化を利用して学習させた．提案手法において， L_{soft} の計算にラベルなしデータを利用しない場合に対応する．
- **KL-regularization (KL)** [Yu 13]: 蒸留による正則化の代わりに KL 正則化を利用して学習させた．提案手法において $T = 1$ とした場合に対応する．また，教師なしデータを利用した半教師あり学習を利用している．

最適化にはすべての手法で Adam [Kingma 15] を利用し，50 エポック学習させた．学習率は 0.0001 とし，その他のパラメタは推奨パラメタとした．また，深いニューラルネットワークを学習する際に汎化能力を高めるために標準的に利用されている Dropout ($p = 0.5$) [Srivastava 14] を利用して学習を行った．他に言及がない限り，提案手法およびその関係手法のハイパーパラメタとしては $T = 10$ および $\lambda = 0.8$ として実験を行い，加えてこれらのパラメタのセンシティブリティに関する分析を行った．

§4 評価方法

提案手法の有効性検証のため，少数のラベル付きデータを持つ新規ユーザに対する認識精度を次の方法により擬似的に評価した．1) 3 人のデータで不特定行動者モデルを学習した．この際，10%を検証用データとして利用することでモデルの評価に利用した．また，検証データを利用した早期終了により学習を停止させた．2) 残りの 1 人のデータのうち，各クラス N_{sup} 個をラベル付き教師データ $\{X'_{sup}, y'_{sup}\}$ として取り出した．また，全データ中 80%をラベルなしデータ $\{X'_{unsup}\}$ として取り出した．3) 半教師ありの手法については $\{X'_{sup}, y'_{sup}\}$ および $\{X'_{unsup}\}$ を，教師ありの手法については $\{X'_{sup}, y'_{sup}\}$ を利用して適応後モデルを学習させた．この際，データが少数となるため検証用データによる早期終了は行わず固定のエポック数だけ学習させた．4) 学習した適応後モデルを学習に利用していない $\{X_{test}\}$ を利用して正解率および平均 F 値により評価した．正解率および平均 F 値は次のとおりである．

$$accuracy = \frac{1}{n} \sum_{i=1}^n 1(\hat{y}_i = y_i),$$

$$Fscore = \frac{1}{|L|} \sum_{l \in L} F^l(\hat{y}, y),$$

ここで， n は全サンプル数， L は行動クラスの集合， y_i と $\hat{y}_i$ は i 番目のサンプルに対する正解ラベルおよび予測ラベル， $F^l(\hat{y}, y)$ は l 番目のクラスに関する F 値を意味する．正解率は全サンプルを均等に扱うのに対して，平均 F 値は少数クラスのサンプルを重要視した指標であり異なる意味を持つ．5) 上記 1) ~ 4) の手順を適応後モデルを作る対象のユーザを変更させて 4 回繰り返した．

4.2 実験結果

提案手法の有用性の検証にあたり，本稿では大別して 3 つの結果について報告する．まず第 1 に，各手法のハイパーパラメタを固定した上での認識精度について比較を行う．次に，提案手法のハイパーパラメタである T および λ を変更した場合の認識精度に対する影響について検証する．これら 2 点により，提案手法である SSD が既存のユーザ適応手法と比較して認識精度の観点で優れていることを定量的に確認する．また，最後に既存の手法と組み合わせた場合の結果について報告し，提案手法がその他の手法と補完的に利用可能であることを示す．

表 1: ラベル付きデータを各クラス N_{sup} 個ユーザ適応に利用した場合の提案手法 (SSD) と既存手法の比較．

(a) $N_{sup} = 5$ (正解率)										
	Ref	Mixin	FT	FT-D	FT-T	SVD	SD	KL	SSD	
S1	0.81	0.01	-0.06	-0.05	-0.06	-0.01	-0.05	0.02	0.02	
S2	0.78	0.02	-0.11	-0.09	-0.11	-0.02	-0.04	0.02	0.02	
S3	0.72	0.04	-0.13	-0.12	-0.14	-0.03	-0.13	-0.00	0.02	
S4	0.78	0.03	-0.15	-0.14	-0.14	-0.05	-0.09	0.04	0.04	
ave	0.77	0.03	-0.11	-0.10	-0.11	-0.03	-0.08	0.02	0.03	
ave ₁₂₄	0.79	0.02	-0.11	-0.10	-0.10	-0.03	-0.06	0.03	0.03	
(b) $N_{sup} = 5$ (平均 F 値)										
	Ref	Mixin	FT	FT-D	FT-T	SVD	SD	KL	SSD	
S1	0.66	0.03	0.00	0.01	-0.00	0.02	-0.00	0.04	0.04	
S2	0.57	0.07	0.01	0.03	0.01	0.03	0.01	0.05	0.05	
S3	0.45	0.17	0.10	0.09	0.10	0.14	0.07	-0.01	0.05	
S4	0.53	0.13	0.09	0.08	0.08	0.11	0.13	0.13	0.13	
ave	0.55	0.10	0.05	0.05	0.05	0.07	0.05	0.05	0.07	
ave ₁₂₄	0.59	0.08	0.03	0.04	0.03	0.05	0.05	0.07	0.07	
(c) $N_{sup} = 20$ (正解率)										
	Ref	Mixin	FT	FT-D	FT-T	SVD	SD	KL	SSD	
S1	0.81	0.02	-0.09	-0.08	-0.10	-0.07	-0.08	0.03	0.05	
S2	0.78	0.05	-0.11	-0.09	-0.10	-0.07	-0.05	0.03	0.06	
S3	0.72	0.07	-0.02	-0.02	-0.01	0.01	-0.04	0.00	0.06	
S4	0.78	0.06	-0.04	-0.04	-0.05	-0.00	-0.02	0.05	0.06	
ave	0.77	0.05	-0.06	-0.06	-0.07	-0.03	-0.05	0.03	0.06	
ave ₁₂₄	0.79	0.04	-0.08	-0.07	-0.08	-0.05	-0.05	0.04	0.06	
(d) $N_{sup} = 20$ (平均 F 値)										
	Ref	Mixin	FT	FT-D	FT-T	SVD	SD	KL	SSD	
S1	0.66	0.08	-0.02	-0.01	-0.03	-0.02	-0.01	0.06	0.10	
S2	0.57	0.17	0.07	0.07	0.06	0.08	0.10	0.08	0.16	
S3	0.45	0.27	0.17	0.18	0.19	0.23	0.17	0.00	0.18	
S4	0.53	0.22	0.15	0.15.15	0.17	0.15	0.19	0.21	0.16	
ave	0.55	0.18	0.09	0.10	0.09	0.11	0.10	0.08	0.16	
ave ₁₂₄	0.59	0.15	0.07	0.07	0.06	0.08	0.08	0.11	0.16	

表 1 に提案手法 (SSD) と比較手法の精度比較結果を示す．表中 S1 ~ S4 は Ref の学習に利用したユーザの ID を示しており，Ref が不特定行動者モデルの認識精度，Mixin から SSD までの各列は Ref からの認識精度の差分を表している（プラスであれば認識精度が改善したことを意味する）．太字+下線 は改善幅がもっとも大きかったことを，太字は改善幅が 2 番目に大きかったことを意味している．また，ave の行は 4 ユーザの平均値を，ave₁₂₄ は参考までに不特定行動者モデルの認識精度が極端に低かったユーザ S3 を除く 3 ユーザの平均値を示している．

結果として，次のような結果が得られた．1. 提案手法は， $N_{sup} = 5$ の場合 $N_{sup} = 20$ の場合双方で正解率および平均 F 値の双方で認識精度に改善が見られた．具体的には， $N_{sup} = 5$ の場合に正解率（平均 F 値）で平均 0.03 (0.07) ポイント， $N_{sup} = 20$ の場合に 0.06 (0.16) ポイント認識精度の改善が見られた．2. 平均的な認識精度において提案手法と同程度の精度を達成したのは，ユーザ適応ではなく各ユーザに対して再学習を行う Mixin のみであった．3. ユーザごとに見ると，ユーザ S3 については，他手法と比較して SSD を用いた場合に精度が悪化する場合があった．例えば， $N_{sup} = 20$ とした場合 Mixin と比較して正解率および平均 F 値の指標でそれぞれ 0.01，0.09 ポイント低くなっていた．

ここで，教師なしデータを利用していない手法群 (表 1 中 FT から SD) のほとんどで平均 F 値が改善している一方で正解率が低下しているのは，全行動クラスのサンプルを均等に N_{sup} 個ずつ与えていることが原因の 1 つと考えられる．このように各行動クラスのサンプルを均等に用意すると，珍しい行動クラスのサンプルが Ref モデルと比較して相対的に強く影響を与えるようになり，結果として珍しい行動クラスのサンプルを識別できるかを主眼とする平均 F 値には良い影響を与える一方で，大多数を占める行動クラスのサンプルを識別できるかを主眼とする正解率は Ref モデルと比較して悪化する．そのため，表 1 が示すところは正解率の指標において特に提案手法で良く改善するというのではなく，正解率および平均 F 値双方の指標で認識精度を改善できていることであることに注意されたい．

次に，図 2 にユーザ適応時の学習エポックごとの平均 F 値の変化 (a が訓練データ，b がテストデータ) を示す．なお，図は $N_{sup} = 20$ とし，ユーザ適応時に元にした不特定行動者モデルの平均認識能力が平均的であった S2 ユーザに対するユーザ適応を行った際の認識精度を描画した．結果として，1)FT や SVD など蒸留による正則化を用いない方法およびラベルなしデータを用いない SD では訓練データにおける認識精度は SSD より高くなっている一方でテストデータにおいては SSD を大きく下回っている，2)KL の場合は SSD と比較すると訓練データに対する認識精度の改善が限定的であり，結果としてテストデータに対する認識精度で SSD の方が高くなって

(a) 訓練データに対する認識精度 (b) テストデータに対する認識精度

図 2: 訓練データとテストデータに対する各手法の学習の推移（横軸がエポック数，縦軸が認識精度）

(a) S1

(b) S3

図 3: $N_{sup} = 20$ のときの Ref と SSD，Mixin の混合行列による比較．(a) は S1，(b) は S3 に関するもの．

いる，という結果が得られた．このことは，ラベルなしデータを利用することで過学習を抑制できており，また T を考慮することでより適切なユーザ適応が行えていることを示している．

図 3 に SSD を行う前後の混合行列および Mixin の混合行列を比較した結果を示す．(a), (b) はそれぞれ Ref の認識精度が最も高かった S1 と最も低かった S3 の結果を示しており，それぞれ左から順に SSD を適用する前の Ref，SSD 適用後，Mixin での混合行列を示している．また，図中色が濃い場所が高い割合であったことを意味しており，0 から 1 まで色を線形にスケールしたものである．結果として，S1 ユーザについては SSD と Mixin の混合行列に大きな差が見られない一方で，S3 ユーザにつ

(a) S1

(b) S3

図 4: 学習に利用した教師なしデータ X_{unsup} とテストデータ X_{test} に対する学習の進行の比較．

図 5: T および λ に関するセンシティビティの分析.

いては例えば誤って 5 番目のクラスと予測してしまうようなエラーが Mixin と比較して SSD で頻出していることがわかる. この傾向は Ref モデルにおいても同様に起こっており, SSD においては適応の際に元のモデルとして利用した Ref のエラーを引きずるという問題が発生していることを示している.

また, 図 4 に SSD の学習に利用した教師なしデータ X_{unsup} と評価用データ X_{test} それぞれに対する学習の進行度合いを SSD と Mixin で比較した結果について示す. 図中青線が SSD, 赤線が Mixin を表しており, 実線が X_{unsup} , 点線が X_{test} に対する認識精度をそれぞれ示している. なお, 図中どこどころ点線が消えているのは実線と重なっているためである. 結果として, まず S3 ユーザについては Mixin を適用した場合に教師なし訓練データとテストデータの認識精度がほとんど同じ値になっている一方で, SSD では, 教師なし訓練データと比べてテストデータの精度が低い値となっていた. また, S1 ユーザについては, 教師なし訓練データとテストデータに対する精度は SSD および Mixin どちらに関してもほとんど同じ精度となっていた. これらの結果から, SSD を S3 ユーザに適用した場合には X_{unsup} にわずかながら過学習しており, 結果として Mixin と比較して認識精度が低くなっていたと考えられる.

図 5 に, ハイパーパラメタ T および λ を変化させた際の認識精度の変化を示している. 横軸は λ を, 縦軸は認識精度を示しており, 各折れ線が異なる T に対応している. 結果として, まず T については, $N_{sup} = 5$ および $N_{sup} = 20$ の場合双方で $T = 1$ の場合 (KL 正則化に対応) と比較して, $T = 3$ や 5, 10 の場合が比較的良好な精度となっていた. この結果は, 提案手法のように温度パラメタを考慮することでより良いユーザ適応が行えていることを示唆している. 次に λ については, T の大きさによるものの, 例えば $T = 10$, $N_{sup} = 5$ の場合には $\lambda = [0.5, 0.9]$ 程度の範囲でおおよそ横ばいであり, $T = 10$, $N_{sup} = 20$ の場合には $\lambda = [0.5, 0.6]$ 付近で最も良い精度となっていた. この結果は, N_{sup} が大きい場合にはより小さな正則化で良いことを示しており, また Yu らが KL 正則化において検証した結果 [Yu 13] とも整合するものである.

図 6 に, 提案手法である SSD と SVD を組み合わせた

図 6: SVD と SSD を組み合わせた場合の認識精度.

図 7: $N_{sup} = 0$ の時の各 T での精度改善度合いの比較.

場合の結果を示す. 横軸は教師ありデータ数 N_{sup} を, 縦軸は平均 F 値を表している. また, k は $W = USV$ により分解した後に S の行列をどの程度利用したかを示しており, 例えば $k = 400$ の場合に特異値が大きい方から上位 400 個を残したことを意味している. 結果として, $N_{sup} = 5$ のように極端に教師ありデータが少ない場合を除くと SVD との組み合わせにより認識精度に改善が見られた. 特に $N_{sup} = 20$ かつ $k = 400$ にした場合に認識精度の改善が大きく, 0.02 ポイント改善が見られた. ただし, SSD と SVD を組み合わせた場合は認識精度の改善により大きなイテレーションを必要としたため, SVD を利用する場合には 200 エポック学習させた結果であることに留意されたい.

最後に, 図 7 にラベルありデータを利用しない ($N_{sup} = 0$) 場合に蒸留によるユーザ適応を行った結果を示す. 図の横軸はユーザを示しており, 棒グラフの各色は T の違いを示している. また, 縦軸は不特定行動者モデルからの認識精度の改善度合いを示しており, 点線が認識精度に全く変化がなかった状態を意味している. 結果から, 教師データが全く無い場合でもユーザ S3 を除く 3 ユーザで認識精度が改善していること, ユーザごとに傾向は異なっているものの平均すると温度 T を高めた方が認識精度の改善度合いが高くなること ($T=20$ の場合に平均して 0.03 ポイントの改善), が確認された.

5. 考 察

まず, 提案手法のユーザ適応による認識精度の観点から見た有効性について考察する表 1 の結果から, ハイパーパラメタおよび固定エポック時点での認識精度を比較した場合, 提案手法が Fine-Tuning や SVD などの既存のユーザ適応手法と比べて平均的に良い精度となることが確認された. また, 図 5 に示した通り提案手法では既存手法と比較して過学習を抑制できており, また学習エポック数の切り出し方によらず安定的に良い精度とであるこ

とがわかる．これらの結果は，提案手法により既存のユーザ適応手法と比較して良いユーザ適応が行えることを示していると考えられる．また，図5の結果は，ほとんどの場合 T を大きくすることで認識精度が向上しており，単に KL 正則化により過学習を抑えるよりも T を考慮することでより良いユーザ適応が行えることを示唆している．表1の結果は，教師なしデータを利用しない SD と比較して教師なしデータを利用した半教師あり学習を行う SSD が優れたユーザ適応を行えることを示している．また，SSD を利用する場合ラベルありデータを全く利用せず蒸留による正則化項のみを利用して学習を行った場合にも認識精度の改善が確認された（図7）．これらの結果から，提案手法はラベルなしデータの情報もうまく活用することで，既存手法と比較して認識精度を高めることができる手法であると言える．

提案手法の認識精度に関する課題として，一部のユーザ（S3 ユーザ）に対して Mixin と比較して認識精度が高まりにくいという問題が確認された．具体的には，図3の結果は提案手法により認識精度は改善するもののユーザ適応の元とするモデルの誤差を引き継ぐことを示していた．提案手法が利用している蒸留による正則化は Ref モデルから離れすぎないように制約をかけることで過学習を抑制する方法であること，また Ref モデルの認識精度が他と比べて著しく低かった S3 ユーザ（平均 F 値で他3 ユーザの平均値より 0.15 ポイント低）に対して提案手法の有効性が相対的に低いことから，元モデルの認識精度が不十分である場合に提案手法の有効性が相対的に低いものと推察される．この問題の解決法としては，1. λ や T を徐々に変化させることで正則化の影響を徐々に小さくする，2. Self-Training と組み合わせることで徐々に教師ではなく自分自身の出力分布に近づける，3. 不特定行動者モデルの学習に利用するユーザ数の増加やアルゴリズム的な工夫により不特定行動者モデルの認識精度をそもそも改善させる，などの方法が考えられる．これらの工夫により認識精度が実際に改善するかの検証が将来課題の1つである．

提案手法に限らずユーザ適応を行う上での認識精度に関する実用上の課題として，ユーザ適応の問題設定での適切なハイパーパラメタの選択方法がある．今回は実験上ハイパーパラメタを固定した上で精度比較を行い，加えてパラメタセンシビリティの検証を行ったが，図5にも示したようにハイパーパラメタの設定は認識精度に影響を与えるため，実用上は何らかの方法でパラメタを選択する必要がある．一般にハイパーパラメタは学習データとは別に用意した教師データ付きの検証用データを用いて行うが，ユーザ適応の問題設定では教師データの数は少なく，ユーザ別の教師データの他に検証用データを各ユーザに対して用意することは現実的には難しい．対策として，例えば今回の実験のように Ref モデルを学習データが大量にあるユーザ全員分で訓練するのではな

く，一部のユーザについては Ref の学習に利用せずユーザ適応の検証ユーザとして利用する方法が考えられる．検証ユーザのデータを利用して実際にユーザ適応を行ったのちにその評価値によりパラメタを決めることによりシステムチックなパラメタ選択が可能である．ただし，この場合検証ユーザに対して良い評価値となるパラメタが別のユーザに対しても良いかどうかは明らかではなく，最適なハイパーパラメタのユーザに関するロバスト性に関する検証が必要と考えられる．

図6では，SSD と既存手法である SVD を組み合わせることにより認識精度が改善することが示された．この結果は，不特定行動者モデルから離れすぎのを防ぐ提案手法と，適応させるパラメタを減らすことで効率良く学習する手法群は補完的であることを示唆している．適応させるパラメタを減らす手法は本稿で利用した一部の層を固定する方法や SVD の他にも，出力層のバイアスを補正する方法 [Yao 12]，ユーザごとに活性化関数の形を変化させる方法 [Siniscalchi 13]，隠れユニットの出力の大きさを補正する方法 [Swietojanski 14, Swietojanski 15] などが音声認識の領域で提案されている．これらの手法の行動認識における有効性，SSD との組み合わせにより認識精度の有効性の調査は更なる精度改善を必要とする実用上の検討事項である．

ユーザ適応を行う上での認識精度以外の観点として，学習モデルを構築するための計算コストがある．今回の実証結果では，ユーザ適応を行わずに各ユーザに対して毎回学習をはじめから行う Mixin が認識精度の上では提案手法と同程度であったが，計算リソースが限られている場合には提案手法のように既に学習済みのモデルが獲得した知識を有効活用できることが望ましい．また，多数のユーザがアプリケーションを利用する場面で認識精度を改善する場面を想定すると，ユーザ適応はサーバ側よりもデバイス側で行えることが望ましい．提案手法において，単純に Fine-Tuning を行う場合と比較して，必要な追加計算はソフトラベル s の計算および L_{soft} の計算のみである．前者についてはニューラルネットワークを順伝搬させる必要があるため全サンプルについて行うにはある程度の計算コストがかかるが，各サンプルについて一度計算すれば良いため実用上そこまで大きな問題にはならない．また， L_{soft} の計算は単純な掛け算であり計算上問題にならないと思われる．DNN のユーザ適応を行う場合に最も計算負荷が大きいのは誤差を逆伝搬させる部分であり，例えばニューラルネットワークの圧縮技術 [Han 16, Lin 16] と組み合わせるなどの改良の余地がある．

本稿の実験設定上の限界および問題点としてまず第1にデータセットに関する限界がある．今回は最もよく利用される Opportunity データセットで検証したが，さらに大規模なデータセット [Zhang 12] やユーザ数の多いデータセットなど多様なデータセットで検証を進めることで

提案手法がどのような場面で有効なのかについての議論を行う必要がある．第 2 に，ユーザ適応時に各クラスに対して均等な個数のサンプル数が与えられるという実験設定に伴う問題がありうる．4.2 節でも述べたように，この問題設定では平均 F 値に対しては良い影響を，逆に正解率に対して悪影響を及ぼす可能性があり，その意味で FT や SVD などの手法を正解率の観点で過小評価している可能性がある．一方で，良い影響を与えられとされる平均 F 値の指標においても，提案手法は FT や SVD を上回っているため，少なくとも今回の問題設定においても提案手法の有効性を示すのに十分であると考えられる．実用上正解率を重要視する場合には，全行動クラスのサンプルを均等に N_{sup} 個ずつ用意するのではなく，Ref モデルの学習時に存在したサンプルの割合に応じて異なる数のサンプルを用意したり，あるいはサンプル数が大きい行動クラスをより強く重み付けした上で学習を行ったりする方法が考えられるが，これらの方法では正解率は高められる一方で平均 F 値は今回行ったような珍しい行動クラスの影響を強く受ける方法と比較して低下するため，実用上はトレードオフを考慮する必要がある．

また，今回の検証ではユーザ適応技術がよく利用されてきた音声認識における研究にならないすべての行動クラスに対して均一な数のラベルデータが得られる仮定のもと検証を行ったが，行動認識の場合実際にはクラスごとにラベルの入手難易度は異なることが多い．例えば，転倒したかどうかを 2 値分類する場合を考えたときに，転倒に関するラベルデータを用意するのは困難であるが転倒していない場合のサンプルを得るのは容易であろう．そのため，すべてのクラスについて同じ数のラベルデータが得られるというよりむしろ特定のクラスのみに対してラベルデータが得られるという仮定の方がより現実に即していると考えられる．Reiss らは，すべての行動クラスに対するラベルデータが得られるという仮定を置くべきではないと主張し，適応データの一部のクラスにのみラベルデータが与えられる場合における検証を行っている [Reiss 13]．提案手法において，温度 T を上げた場合に得られるソフトターゲットはクラス間の類似性を保持した情報となるため，特定のクラスにしかラベルがない場合でも良い精度が出るのではないかと期待できる．

6. お わ り に

Deep Neural Networks は行動認識の分野においても高い認識精度を達成することが近年の研究で示されており，人間行動を高精度に認識する上で重要な技術となっている．本稿では，学習済みの深層行動認識モデルをユーザ適応する問題に対して，蒸留を応用した方法を提案した．ユーザ適応はラベル付きデータが限られたユーザに対する認識精度を改善する主要な方法の 1 つであるが，深層行動認識モデルをいかにユーザ適応させるかの検討はこ

れまで行われてこなかった．DNN をユーザ適応する際の最大の論点は，ユーザ適応に利用可能なラベル付きデータは大量のパラメータを持つ DNN 全体を適切に訓練するほど十分に得られないことが多く，単純に適応させようとすると過学習を引き起こす点である．提案手法は，蒸留をユーザ適応の文脈およびにラベルなしデータを活用できるように拡張し，ユーザ適応後のモデルが元のモデルから極端に離れないように制約をかけることで過学習を抑制してユーザ適応する方法である．本稿では実データによる実験を行い，提案手法により 8 種類のベースライン手法と比較してより良いユーザ適応が行えることを示した．今後は，より大規模なデータセットでの検証やより実環境に近い問題設定において検証を行うことで，現実的な環境での行動認識の高精度化に貢献していきたい．

◇ 参 考 文 献 ◇

- [Chapelle 05] Chapelle, O. and Zien, A.: Semi-supervised classification by low density separation, in *AISTATS*, pp. 57–64 (2005)
- [Dehak 11] Dehak, N., Kenny, P. J., Dehak, R., Dumouchel, P., and Ouellet, P.: Front-end factor analysis for speaker verification, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 19, No. 4, pp. 788–798 (2011)
- [Grandvalet 04] Grandvalet, Y. and Bengio, Y.: Semi-supervised learning by entropy minimization, in *Advances in Neural Information Processing Systems (NIPS)*, pp. 529–536 (2004)
- [Hammerla 16] Hammerla, N. Y., Halloran, S., and Ploetz, T.: Deep, convolutional, and recurrent models for human activity recognition using wearables, *International Joint Conference on Artificial intelligence (IJCAI)* (2016)
- [Han 16] Han, S., Mao, H., and Dally, W. J.: Deep compression: Compressing deep neural network with pruning, trained quantization and huffman coding, in *International Conference on Learning Representation (ICLR)* (2016)
- [Hinton 15] Hinton, G., Vinyals, O., and Dean, J.: Distilling the knowledge in a neural network, *arXiv preprint arXiv:1503.02531* (2015)
- [Ioffe 15] Ioffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift, *arXiv preprint arXiv:1502.03167* (2015)
- [Karanasou 14] Karanasou, P., Wang, Y., Gales, M. J. F., and Woodland, P. C.: Adaptation of deep neural network acoustic models using factorised i-vectors., in *INTERSPEECH*, pp. 2180–2184 (2014)
- [Kingma 15] Kingma, D. and Ba, J.: Adam: A method for stochastic optimization, in *International Conference on Learning Representation (ICLR)* (2015)
- [Lane 11] Lane, N. D., Xu, Y., Lu, H., Hu, S., Choudhury, T., Campbell, A. T., and Zhao, F.: Enabling large-scale human activity inference on smartphones using community similarity networks (CSN), in *International Conference on Ubiquitous Computing (UbiComp)*, p. 355, New York, USA (2011)
- [Lee 13] Lee, D.-H.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks, in *Workshop on Challenges in Representation Learning (ICML)*, Vol. 3, p. 2 (2013)
- [Lin 16] Lin, Z., Courbariaux, M., Memisevic, R., and Bengio, Y.: Neural networks with few multiplications, in *International Conference on Learning Representation (ICLR)* (2016)
- [Lopez-Paz 16] Lopez-Paz, D., Bottou, L., Schölkopf, B., and Vapnik, V.: Unifying distillation and privileged information, in *International Conference on Learning Representation (ICLR)* (2016)
- [Ordóñez 16] Ordóñez, F. J. and Roggen, D.: Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition, *Sensors*, Vol. 16, No. 1, p. 115 (2016)

- [Plötz 11] Plötz, T., Hammerla, N. Y., and Olivier, P.: Feature learning for activity recognition in ubiquitous computing, in *International Joint Conference on Artificial Intelligence (IJCAI)*, p. 1729 (2011)
- [Reiss 13] Reiss, A. and Stricker, D.: Personalized mobile physical activity recognition, in *International Symposium on Wearable Computers (ISWC)*, pp. 25–28 (2013)
- [Rifai 11] Rifai, S., Dauphin, Y. N., Vincent, P., Bengio, Y., and Muller, X.: The manifold tangent classifier, in *Advances in Neural Information Processing Systems (NIPS)*, pp. 2294–2302 (2011)
- [Roggen 10] Roggen, D., Calatroni, A., Rossi, M., Holleczeck, T., Förster, K., Tröster, G., Lukowicz, P., Bannach, D., Pirkel, G., Fersch, A., and al, et.: Collecting complex activity datasets in highly rich networked sensor environments, in *Networked Sensing Systems (INSS)*, pp. 233–240 (2010)
- [Romero 15] Romero, A., Ballas, N., Kahou, S. E., Chassang, A., Gatta, C., and Bengio, Y.: Fitnets: Hints for thin deep nets, in *International Conference on Learning Representation (ICLR)* (2015)
- [Sagha 11] Sagha, H., Digumarti, S. T., Millán, J. d. R., Chavarriaga, R., Calatroni, A., Gen, D., and Tröster, G.: Benchmarking classification techniques using the opportunity human activity dataset, in *Networked Sensing Systems (INSS)*, pp. 36–40 (2011)
- [Saon 13] Saon, G., Soltan, H., Nahamoo, D., and Picheny, M.: Speaker adaptation of neural network acoustic models using i-vectors, in *Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp. 55–59 (2013)
- [Siniscalchi 13] Siniscalchi, S. M., Li, J., and Lee, C.-H.: Hermitian polynomial for speaker adaptation of connectionist speech recognition systems, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 21, No. 10, pp. 2152–2161 (2013)
- [Srivastava 14] Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: A simple way to prevent neural networks from overfitting, *Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 1929–1958 (2014)
- [Swietojanski 14] Swietojanski, P. and Renals, S.: Learning hidden unit contributions for unsupervised speaker adaptation of neural network acoustic models, in *IEEE Spoken Language Technology Workshop (SLT)*, pp. 171–176 (2014)
- [Swietojanski 15] Swietojanski, P. and Renals, S.: Differentiable pooling for unsupervised speaker adaptation, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4305–4309 (2015)
- [Vollmer 13] Vollmer, C., Gross, H.-M., and Eggert, J. P.: Learning features for activity recognition with shift-invariant sparse coding, in *International Conference on Artificial Neural Networks*, pp. 367–374, Springer (2013)
- [Weiss 12] Weiss, G. M. and Lockhart, J. W.: The impact of personalization on smartphone-based activity recognition, in *AAAI Workshop on Activity Context Representation: Techniques and Languages* (2012)
- [Weston 12] Weston, J., Ratle, F., Mobahi, H., and Collobert, R.: Deep learning via semi-supervised embedding, in *Neural Networks: Tricks of the Trade*, pp. 639–655, Springer (2012)
- [Xue 14] Xue, J., Li, J., Yu, D., Seltzer, M., and Gong, Y.: Singular value decomposition based low-footprint speaker adaptation and personalization for deep neural network, in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 6359–6363 (2014)
- [Yang 15] Yang, J. B., Nguyen, M. N., San, P. P., Li, X. L., and Krishnaswamy, S.: Deep convolutional neural networks on multichannel time series for human activity recognition, in *International Joint Conference on Artificial Intelligence (IJCAI)* (2015)
- [Yao 12] Yao, K., Yu, D., Seide, F., Su, H., Deng, L., and Gong, Y.: Adaptation of context-dependent deep neural networks for automatic speech recognition, in *IEEE Spoken Language Technology Workshop (SLT)*, pp. 366–369 (2012)
- [Yu 13] Yu, D., Yao, K., Su, H., Li, G., and Seide, F.: KL-divergence regularized deep neural network adaptation for improved large vocabulary speech recognition, in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 7893–7897 (2013)
- [Zhang 12] Zhang, M. and Sawchuk, A. A.: USC-HAD: A daily

activity dataset for ubiquitous activity recognition using wearable sensors, in *ACM International Conference on Ubiquitous Computing (Ubicomp) Workshop on Situation, Activity and Goal Awareness (SAGAware)*, Pittsburgh, Pennsylvania, USA (2012)

- [Zhao 11] Zhao, Z., Chen, Y., Liu, J., Shen, Z., and Liu, M.: Cross-person mobile-phone based activity recognition, in *International Joint Conference on Artificial Intelligence (IJCAI)*, Vol. 11, pp. 2250–2254 (2011)

〔担当委員：寺島 裕貴〕

2016 年 8 月 29 日 受理

著者紹介

岩澤有祐

2012 年に上智大学理工学部卒業，2014 年に同大学院理工学系研究科修士課程，2017 年に東京大学大学院工学系研究科博士課程修了．博士（工学）．2017 年より東京大学大学院工学系研究科技術経営戦略学専攻 松尾研究室 特任研究員．専門はウェアラブルセンサによる行動認識，障害者支援，機械学習・深層学習応用．

矢入郁子 (正会員)

1994 年東京大学工学部卒業．1996 年同大学院工学系研究科修士課程，1999 年同博士課程修了．同年，郵政省通信総合研究所（現 情報通信研究機構）に研究官として入所，2008 年上智大学に転職，現在，理工学部准教授．2006 年ドコモ・モバイル・サイエンス賞社会科学部門奨励賞を受賞．2012～2016 年 HI 学会理事，その間 2012～2014 年度同学会学会誌委員会委員長．ACM，情報処理学会，日本機械学会会員．博士（工学）．

松尾豊 (正会員)

1997 年東京大学工学部卒業．2002 年同大学院博士課程修了．博士（工学）．産業技術総合研究所，スタンフォード大学を経て，2007 年より，東京大学大学院工学系研究科技術経営戦略学専攻 准教授．2012 年より本学会理事・編集委員長，2014 年より倫理委員長．人工知能学会論文賞，情報処理学会会長尾真記念特別賞，ドコモ・モバイル・サイエンス賞などを受賞．専門は，Web 工学，Deep Learning，人工知能．

◇ 付 録 ◇

A. ニューラルネットワークの構造

図 A.1: 実験に利用した畳み込みネットワークの構造．図中黄色枠が入力，紫が出力，緑が畳み込み層，青がプーリング層，赤が全結合層．また，畳み込み層およびプーリング層のとなりにある $1 \times a$ という表記は畳み込みまたはプーリングの大きさを，各層の右下にある数値は特徴マップ数あるいはユニット数を意味する．