博士論文

ウェアラブルセンサを利用した人間行動の計測のための表現学習に関する研究

指導教員　松尾 豊 准教授

東京大学大学院 工学系研究科

技術経営戦略学専攻

岩澤 有祐

目 次

第 1 章　序論 **1**

1.1　背景 1

1.1.1　特徴設計と表現学習 3

1.1.2　ウェアラブルセンサを利用した行動認識と表現学習 5

1.2　本研究の目的 7

1.3　本論文の構成 8

第 2 章　ウェアラブルセンサを利用した行動認識に関する既存研究 **11**

2.1　行動認識の類型化とウェアラブルセンサの位置づけ 11

2.2　ウェアラブルを利用した行動認識モデルの構築方法 15

2.2.1　センサによる行動の観測 16

2.2.2　セグメンテーション 18

2.2.3　特徴量への変換 18

2.2.4　分類器の学習 21

2.3　ウェアラブルを利用した行動認識モデルの評価 22

2.3.1　ユーザ独立認識精度とユーザ依存認識精度 22

2.3.2　プライバシーへの配慮 27

第 3 章　ウェアラブルセンサを利用した行動認識のための表現学習 **31**

3.1　表現学習と深層学習 31

3.2　ウェアラブルセンサを利用した行動認識に表現学習を適用した既存研究 35

3.3　本研究の論点と 4 章以降の位置づけ 37

3.4　ニューラルネットに関する前提知識 40

3.4.1　順伝播型ニューラルネット 40

3.4.2　ディープネット学習のための工夫 44

3.4.3　畳み込みニューラルネット 48

3.4.4　蒸留 53

第 4 章　ディープネットによるユーザ独立認識と学習された表現の分析 **55**
4.1　ユーザ独立認識とディープネット 55
4.2　利用データセット・ネットワーク構造 57
4.2.1　データセット 57
4.2.2　ニューラルネットの構造 58
4.2.3　比較対象 59
4.3　評価方法 60
4.3.1　前処理 61
4.3.2　ユーザ独立認識性能の評価方法 61
4.3.3　学習された表現のクラスタリングによる評価 62
4.4　実験結果 63
4.4.1　CNN と経験的特徴量の比較 63
4.4.2　Residual Networks による分析 66
4.4.3　行動クラスごとの分析 67
4.4.4　特徴量のユーザ依存性の定性評価 68
4.5　関連研究・考察 69

第 5 章　ユーザ敵対的ニューラルネット: 敵対的訓練によるユーザ独立な表現学習法 **73**
5.1　表現のユーザ独立性 73
5.2　関連研究 75
5.3　ユーザ敵対的ニューラルネット 76
5.3.1　定式化 76
5.3.2　UANN の構造 76
5.3.3　$\mathrm{L}_y(X, Y)$ の定量化 76
5.3.4　$\mathrm{L}_u(R, U)$ の定量化 77
5.3.5　最適化 78
5.3.6　相互情報量最小化との関係 79
5.4　実験 80
5.4.1　データセット 80
5.4.2　実験設定 81
5.4.3　評価方法 82
5.4.4　実験結果 84

5.5 考察 88
5.5.1 ユーザ独立性について 88
5.5.2 ユーザ汎化性能について 89
5.5.3 プライバシー保護について 90

第 6 章 半教師あり蒸留による深層行動認識モデルのユーザ適応 92
6.1 ユーザ適応 92
6.2 関連研究 94
6.2.1 行動認識でのディープネットの利用 94
6.2.2 深層行動認識モデルのユーザ適応 95
6.2.3 蒸留によるユーザ適応 96
6.3 半教師あり蒸留によるユーザ適応 98
6.3.1 アルゴリズム 98
6.3.2 既存手法との関係 99
6.4 実験 101
6.4.1 実験設定 101
6.4.2 実験結果 103
6.5 考察 108

第 7 章 考察 113
7.1 4 章から 6 章の結果まとめ 113
7.2 本論文の貢献と課題 114
7.2.1 正確な行動認識に向けた貢献と課題 114
7.2.2 プライバシーへの配慮に対する本論文の貢献と課題 119
7.3 ウェアラブルセンサによる行動認識技術の活用可能性とその課題 . . . 122
7.3.1 ウェアラブルセンサを取り巻く環境と将来の展望 122
7.3.2 ウェアラブルセンサによる行動認識の活用に向けた課題 . . . 124

第 8 章 結論 127

参考文献 129

付 録 A 車いすセンシングによる路面アクセシビリティ情報収集システム 148
A.1 システム概要 148

A.2 ウェアラブルセンサを利用した行動認識に関するデータセットと転移アプローチの難しさ 151
A.3 車いす屋外行動データセットの収集 155
A.3.1 センシングシステム 155
A.3.2 実験参加者 156
A.3.3 実験方法 156
A.3.4 アノテーション 156
A.4 評価実験 157
A.4.1 実験 1：経験的特徴量を利用した基礎検討 157
A.4.2 実験 2：データセットと認識精度に関する検討 163

付 録 B 5 章に関する付録 168
B.1 交差エントロピー最大化と相互情報量 168
B.2 複数のウィンドウを利用した場合のユーザ予測精度に関する検討 . . . 169

付 録 C 発表業績一覧 170

図 目 次

2.1 ウェアラブルセンサを利用した行動認識モデル構築の一般的な流れ. 16
2.2 特徴量のユーザ依存性がもたらす問題. 赤，青，緑がそれぞれ異なるユーザを，ダイアモンド型のサンプルが行動 1，丸型のサンプルが行動 2，また点線が赤青ユーザに対して 100%の精度で認識可能な超平面の 1 例を示している. 25
2.3 特徴量のユーザ独立性がもたらす利点の概念図. 25

3.1 本論文の論点と以降の章の対応関係. 38
3.2 順伝播型ニューラルネットの基本的な構成要素. 41
3.3 3 層構造を持つ順伝播型ニューラルネットの模式図. 42
3.4 ドロップアウトの概念図. 図中点線で描かれた丸は擬似的に削除されたユニットを指す. 46
3.5 全結合層と畳み込み層の層間の結合方法の比較. 図中の黒く塗りつぶされた箇所はある入力ユニットがどの範囲に影響をあたえるかを示したものである. 49
3.6 プーリングによる移動不変性の概念図. 各図中左側が畳み込み処理による出力ユニットを，右側が最大プーリングの結果を示したものである. 51

4.1 分析に利用した Residual Networks の概要，(a)Residual Block の詳細，(b)Residual Block を利用したネットワーク構造の全体像. 59
4.2 ディープネットの学習および評価の流れ. 60
4.3 最大ユーザ傾向の計算方法. 62
4.4 経験的な特徴量の追加による認識精度の変化度合いに関するヒストグラム. 66
4.5 $\mathrm{Res}_{\mathbf{L}}$（$L$=6, 12, 18, 24）を利用した場合の認識精度の比較. 67
4.6 テストデータと検証データでの混同行列の差分の可視化. 69

4.7 クラスタ数を20とした際の各クラスタのユーザ傾向の可視化．(a)Opportunity, (b)Heterogeneous 70
4.8 クラスタ数を変化させた際の各層におけるMUTの比較 71

5.1 ユーザ敵対的ニューラルネットの構造と一般的なMLPの比較． . . . 77
5.2 実験に利用した提案手法のネットワーク構造． 80
5.3 行動認識精度とユーザ独立性のバランスの比較．各マーカはテストユーザ，各色は手法の違いを表している．また，ダイヤモンドマーカは全ユーザの平均値を表す． 83
5.4 行動認識精度およびユーザ認識精度のエポックごとの変化． 85
5.5 クラスタ数を20とした際のOpp-Gにおける各クラスタのユーザ傾向の可視化．a: CNN，b: UANN (λ=0.1)． 88

6.1 ユーザ適応の概念図． 93
6.2 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） 106
6.3 $N_{sup} = 20$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． 107
6.4 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． 108
6.5 Tおよびにλに関するセンシティビティの分析． 109
6.6 SVDとSSDを組み合わせた場合の認識精度． 109
6.7 $N_{sup} = 0$の時の各Tでの精度改善度合いの比較． 110

A.1 車いすセンシングによるアクセシビリティ情報収集システムの概念図．149
A.2 実験に利用した車いす行動のセンシングシステムおよびに取得実験の様子． 156
A.3 走行経路に関する情報．(a) 経路概要, (b) 平らな路面, (c) タイル路面, (d) 坂（傾度5度以上）, (e) 車道と歩道の間の縁石, (f) タイル路面の間の小さな段差, (g) 木の根により盛り上がったコンクリート路面． 157
A.4 特徴量と認識精度の関係の比較． 160
A.5 分類器の比較． 160
A.6 訓練データ数と認識精度の関係． 161

A.7 ウィンドウサイズがF値に与える影響. 161
A.8 27周分の推定結果の可視化結果．縁石，点字ブロック，坂をそれぞれ赤，黄色，緑色の円で可視化した. 162
A.9 非個人モデルの学習に利用するユーザ数を増加させた際の認識精度の変化．各行がそれぞれ各特徴量に，また左列が検証データに対する認識精度，右列がテストデータに対する認識精度を示している. 166
A.10 学習に利用するデータ量に対する認識精度の関係. 167

B.1 複数のウィンドウでの出力確率の平均によるユーザ識別性能の変化. 169

表 目 次

4.1 分析に利用したデータセットのメタ情報. 61
4.2 CNN と既存の特徴量の認識精度における比較. 65
4.3 行動クラスごとの認識精度. 68

5.1 行動認識精度に関する比較. 84
5.2 異なるユーザ分類器を利用した場合のユーザ認識精度の比較. 86

6.1 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. 104

A.1 学習に利用するユーザ数を増やした際に認識精度が改善したも割合の比較. 164

第1章　序論

1.1　背景

近年，小型かつ多様なセンサを搭載したリストバンド型センサや腕時計型センサなど多種多様なウェアラブルデバイスが登場し，一般向けに普及し始めている．また，多くのユーザが持ち歩くスマートフォンは加速度センサやジャイロスコープなどユーザの状態を識別するのに有効なセンサを搭載している．このようなユーザが身に着けて持ち歩くウェアラブルセンサの普及に伴い，実世界の行動計測可能性は大きく高まった．実世界のユーザの行動を正しく認識し蓄積することができれば，ヘルスケアやコンテキストアウェアをはじめとしてさまざまなアプリケーションが実現可能になり意義深い．

ウェアラブルセンサを利用して人間行動を計測する技術の最も代表的な応用先はヘルスケア領域である．すでに実世界で応用されている事例としては，万歩計や運動量計，歩行アドバイザなどがある．例えば，Jowbone Up[1] 及びにその関連アプリケーションではユーザの活動量や睡眠状態を計測したり，運動のアドバイスを行ったりしている．また，より高度な認識を利用したシステム開発を目指すプロジェクトの1つに，モバイルデバイスによるセンシングを利用したヘルスケアシステムの構築を目指す m-Health (mobile-Healthcare) がある [1]．m-Health とは Mobile Healthcare の略語であり，ユーザがもちあるくモバイルデバイスをユーザの状態をセンシングすることで，例えば医者が患者の健康状態を遠隔から確認したり，あるいは患者自身が自身の健康状態を確認したりといったシステム実現による予防医療の深化が期待されている．

より具体的な研究事例として，例えば MIT で開発された LiveNet は，3 軸加速度センサ，EEG，EMG などのセンサを利用してパーキンソン病患者のモニタリングやてんかん性発作 (epileptic seizures) の検出を試みている [2]．Steele らは，慢性閉塞性肺疾患 (chronic pulmonary disease) 患者の行動をセンサから認識することで，リハビリテーションの補助や研究による病理理解を補助することを提案している [3]．

[1] https://jawbone.com/

あるいは，肥満患者の計測した行動履歴を元にライフスタイルや行動の改善など治療の補助に利用するシステムも提案されている [4, 5]．身体的なデータのみならず，Ertin らは，音声センサや加速度センサ，心拍センサなどを利用してユーザのモニタリングを行うことで，そのストレス状態を 90%以上の精度で推定可能であることを示している [6]．

ヘルスケアとは別の応用領域として，コンテキストアウェアシステム (context-aware system) と呼ばれる領域もある．コンテキストアウェアシステムとは，ユーザの行動や状態を認識し理解することで周辺のデバイスを自動的に制御することでユーザの快適性を向上したりユーザの行動を支援するシステムの総称である [7]．ユーザが常に身に着けているデバイスがユーザの行動を計測し，行動認識を行っていれば，例えば席を立って部屋を出るだけで不要な電灯や空調を切ったり，エレベータを呼び出したりといったことが可能になる．あるいは，ジョギング時にはスマートフォンの Wi-Fi をオフにすることでバッテリー消費を抑制することもできるようになる．

また，単に認識対象のユーザの行動を知りたい場合だけでなく，実世界で人間がどのような行動をしているのかをデジタル化し記録・解析することで新たな知見を得ようとする試みもある．例えば日々の行動と病歴などを突合することができれば，どのような行動をしている人がどの病気になりやすいのか，どのような行動によりそれを防ぐことができるのかなどを情報を精緻に収集することが可能になりうる．あるいは，看護師の病院内行動データを収集することで，例えば看護師が次にどこに行き何を行うべきかを提示することでより効率的な業務遂行のサポートをすることも可能になる [8]．また，われわれは車いすユーザの行動を計測し蓄積することで，路面の状態や安全性を推定するシステムを提案している [9]．

これらのようにユーザの持ちあるくセンサにより人間行動を計測することは多様な応用先を持ちその利用のされ方はアプリケーションにより異なる．一方，センサデータはそのままではノイジーで扱いにくく，機械学習を始めとした何らかの統計的処理により観測したデータを行動やコンテキストに変換するという共通の技術的側面を有している．本論文が扱うウェアラブルセンサによる行動認識 (wearable-sensor based activity recognition) は，これらのようにユーザが身に付けたデバイスにより観測されたセンサ値を利用してそのユーザの行動やその状態，コンテキストを認識することを目指す研究領域である．ウェアラブルセンサによる行動認識の高精度化などにより，実世界の人間行動の観測を基礎とするシステムの創発が期待されている．

ウェアラブルデバイスによる観測値からユーザの行動を予想するという問題は，機械学習の観点から捉えると何らかの観測値 X から予測したい情報 y を推定する関数

f を学習する問題の 1 例として捉えることができる．この時，多様な X に対して正しい y を予測できるような関数 f を学習することが目的となる．このような良い写像を学習する技術として近年重要度を増している技術の 1 つに表現学習 (representation learning) がある [10]．詳しくは，1.1.1 項で説明するが，表現学習は高次元の観測信号 X を情報 y を予測しやすい表現（あるいは特徴量）r に変換する技術であり，画像認識や音声認識など高次元の生データを入力にとる認識問題を高精度に行う上で有効であることが報告されている．以降 1.1.1 項では表現学習について説明し，1.1.2 項で表現学習の代表的な応用先である画像認識や音声認識と比較した場合の行動認識の特徴を整理することで，表現学習をウェアラブルセンサによる行動認識に適用する場合の論点について整理する．

1.1.1 特徴設計と表現学習

1.1 節においてウェアラブルセンサを利用した行動認識問題をセンサによる観測値 X から認識したい行動 y への対応を学習する問題であると記述した．しかし，観測 X が画像や音声，あるいはウェアラブルセンサによる行動認識において利用される加速度センサのように高次元なものである場合観測値 X から直接 y への対応を学習することは特に Support Vector Machine (SVM) や決定木などの従来的な機械学習の手法で行うことは難しい場合が多い．そのため，観測信号 X を人手で設計した特徴抽出器により特徴量 r と呼ばれる値に変換した後に，特徴量と y の対応関係を学習するという方法が取られてきた．Bengio らも指摘するように [10]，機械学習を利用する場合学習器がうまく識別できる特徴抽出器を人手で設計することが認識精度に大きな影響を与えており，例えば入力が画像であれば SURF や HOG，音声認識であれば MFCC，ウェアラブルセンサによる行動認識では後に説明する ECDF (Empirical Cumrative Distribution Function) [11] など，領域に特化した特徴量設計手法の開発により認識精度を高める試みがなされてきた．以降便宜上，このような人手による特徴抽出器を設計する方法を特徴設計と呼ぶこととする．

このような特徴設計は認識精度を高める上で重要な役割を持つ一方で，ウェアラブルセンサによる行動認識において有効な特徴量の設計はいくつかの理由から難しく，利用されている特徴量による限界が指摘されてきた [12, 11]．ウェアラブルセンサによる行動認識に有効な特徴量を設計する第 1 の難しさは，観測される信号の解釈困難性である．ウェアラブルセンサを利用した行動認識でよく利用されるセンサには加速度センサやジャイロセンサなどの慣性センサがあるが，これらは通常人間

が目で確認しても解釈不可能である．そのため，近年では ECDF [11] などの手法が提案され始めているものの，基本的には平均値や相関係数など単純な統計量が利用されることがほとんどであり認識精度の観点でのボトルネックとなっている．

第 2 の難しさは，認識したい対象の多様性である．例えば，肥満者のフィットネス管理目的で利用するには寝ている時間，歩いている時間，走っている時間などを認識できれば良いかもしれないが，パーキンソン病の進行度合いを確認する目的で利用する場合にはパーキンソン病特有の行動を識別できる必要がある．また，一般に対象とする行動が何かによって有効な特徴量は異なる．例えば，歩く/走る/座ると言った基本的な行動を認識したい場合には信号の強さなどの統計的な値で十分認識できる可能性が高いが，ドロワーを開ける，包丁で切るなどのキッチンにおける行動を認識したい場合にはセンサのパターンなどより精緻な特徴量が有効となるだろう．このように，アプリケーションごとに最適な特徴量を選択したり設計したりすることが認識精度を高める上で重要な要素であるものの現実的にはコストが嵩んでしまうため難しい．

データをどう表現するのかを人間の探索によって決める特徴設計というアプローチに対して，表現学習とはデータから表現を獲得するというアプローチである [10]．表現学習はディープネットワーク (deep neural networks, 以降ディープネット) という階層性という仮定を考慮することにより効率的に表現を学習するアルゴリズムの普及を背景に立ち上がった新興領域であり，画像認識 [13, 14, 15]，音声認識 [16]，ゲームエージェント [17]，コンピュータ囲碁 [18][2] などさまざまな領域を著しく発展させている．

ディープネットの主要な応用先は画像認識や音声認識といった分野であるものの，本研究が対象とするウェアラブルセンサによる行動認識の領域においても高い精度で行動を認識する上での有効性が認識され始めている．表現学習がウェアラブルセンサを利用した行動認識に適していると考えられている理由は大別して次の 2 つである．

第 1 に，従来の特徴設計は複雑な特徴の利用が難しいという性質を持っていた．つまり，人間が信号の観察により思いつくことのできる特徴量は比較的単純なものであるか，あるいは仮に信号の中に複雑な規則を見出すことができたとしてもその規則を数値化することは困難である．一方で，表現学習では何らかの損失関数を最小化する表現を自動で発見するため，その意味で人間の能力の限界により利用できる特徴量が限定されない．第 2 に，従来の特徴設計は人間がそのプロセスに必要不可

[2] 2016 年 3 月には当時のトッププロであるイ・セドルを破った．

欠であるため自動化が難しいのに対して，表現学習では損失関数を最小化する問題として機械的に特徴を獲得する．そのため，アプリケーションごとに適切な特徴量を機械的に学習できる蓋然性があり，多様な認識対象が考えうるウェアラブルセンサによる行動認識においても各アプリケーションごとに適切な特徴設計を比較的低コストに行えることが期待できる．

これらの背景から，ディープネットを始めとした表現学習について，2011 年に Plotz らが初めて同種のタスクに適用した [12] のを皮切りにモデル構造に関する比較などが進められてきている．以降本論文でも表現学習が行動認識において高い認識精度を達成する上で重要な技術であると捉え論を進める．

1.1.2 ウェアラブルセンサを利用した行動認識と表現学習

1.1.1 項に示した通り，データから適切な表現を獲得することを試みる表現学習は高次元の人間が解釈しにくいデータから知識を獲得することを試みるウェアラブルセンサを利用した行動認識とも親和性が高く，認識精度を高める上で重要な要素として認識され始めている．一方，表現学習は画像認識や音声認識を中心に研究されてきた技術であり，ウェアラブルセンサによる行動認識に対して実用的に利用する上ではいくつかの不都合な点がある．本項では，既存の表現学習技術の主要な応用先である画像認識や音声認識の問題とウェアラブルセンサによる行動認識を比較することによりウェアラブルセンサによる行動認識のタスクの性質について説明し，表現学習をウェアラブルセンサによる行動認識に適用する際の論点について整理する．

画像認識や音声認識とのウェアラブルセンサによる行動認識を比較した場合の第 1 の違いに，扱うデータがパーソナルデータであるという点がある．ウェアラブルセンサによる観測されるデータは，ユーザが持ちあるくセンサによる観測値であるという性質上特定のユーザと強く紐づくデータである．このようなパーソナルデータを扱う行動認識に対して表現学習を利用するうえでは認識精度を高めるという観点だけでなく，いかにユーザのプライバシーに配慮するという観点が必要となる．

このようなパーソナルデータを扱う場合にプライバシー対策を行うことは，単にユーザの気持ち悪さを排除するという以上の意味合いがある．例えば，データが適切に管理されているかわからなかればユーザはそのアプリケーションを利用しないであろうし，予期しない情報が漏えいすることになればユーザはさまざまな実害を被りうる．またアプリケーション提供者の観点から見ても，プライバシーに配慮されていない大規模なデータを保持することはプライバシー漏えいリスクを抱えるこ

とに繋がるため望ましくない．一般にユーザの情報が漏えいすることになった場合の企業価値の損失は甚大であるから，プライバシーに配慮することはユーザが安心して参画できるようにする，またアプリケーション提供者が大量データを集めて質の高いアプリケーションを開発・提供できるようにする上で重要な観点となる．

ウェアラブルセンサによるデータから表現を学習する場合の2つ目の問題点として，ラベルデータが少ないユーザに対して有用な表現をいかに学習するかという問題がある．一般にディープネットによる高い認識精度を引き出すためには，大規模な教師付きデータセットを利用する必要がある．しかし，ウェアラブルセンサによる行動認識で代表的に利用される加速度センサやジャイロセンサ，あるいは心拍センサなどは人間が目で見て事後にラベル付けすることは困難である．これは，画像は人が目で見てラベル付けを行うことで，あるいは音声は人が聞いてラベル付け可能であることと対照的である．

ウェアラブルセンサを利用した行動認識においては一般に，ラベル付きデータは次のいずれかの方法で収集される．第1の方法は，アプリケーション開発者などが用意した実験環境に参加したユーザの行動を動画などで記録することにより，ユーザの行動と行動のラベルを対応付ける方法である．動画データを外部データとして利用するため信頼度の高いラベルデータを作成することができ，Opportunity Challenge Dataset，USC-HAD，Heterogeneous Activity Recognition Dataset など公開された行動認識に関するデータセットの多くはこの方法で作成されている．第2の方法は，ユーザにこれから取る行動，あるいは直前に取った行動を入力してもらう形でラベルを付与する方法である．第1の方法と比較すると低コストでラベルデータを取得できる可能性がある反面，各ユーザの負担が大きくなったり，ラベルデータの信頼度が低くなったりする可能性がある．公開されたデータセットでは，例えば ALKAN Dataset はこの方法で作成されたデータセットである．

いずれの方法を利用する場合でも，認識対象とするユーザそれぞれに対してラベル付きデータを大量に取得することはコストが嵩んだりユーザの負担が大きくなるため現実的には難しいと考えられる．また，ユーザの負担という観点の他に，行動認識では対象とする行動として転倒行動や病気に特有の行動などそもそも発生頻度が低いアブノーマルな行動を対象とすることもラベルデータ取得の難しさの1つの要因である．例えば転倒行動を検出する問題を考えた場合に，転倒という行動を多くのユーザに対して付与することは危険であり，また発生頻度が低く大量のデータを得ることは困難である．これらの理由から，ウェアラブルセンサによる行動認識を対象とする場合，教師付きの学習データが全くないあるいは少ないユーザに対し

ても良い表現を学習できるような仕組みがウェアラブルセンサによる行動認識に対して表現学習を利用する上では必要となる．

1.2 本研究の目的

本研究は，ウェアラブルセンサにより得られるデータから正確に人間行動を認識する上でディープネットが重要になりつつあることを背景に，ウェアラブルセンサによる行動認識に対して有効な表現学習技術の開発を目的とするものである．具体的には，表現学習がよく利用されてきた画像認識や音声認識と比較してウェアラブルセンサを利用した行動認識のタスクの性質上問題となるプライバシーへの配慮およびに教師付きデータの少ないユーザに対する認識精度を高めるという観点から手法の提案を行うことで，ウェアラブルセンサによる行動認識に対して表現学習技術およびにその拡張手法が適用可能であることを示すものである．

ウェアラブルセンサによる行動認識のための表現学習技術を提案するという目的に対して，次の 3 つのサブ目的を取り扱う．

- ディープネットが行動認識精度を高めるために有効であることを示すこと
- プライバシーに配慮した表現を学習する方法を提案すること
- 教師付きデータが少ないユーザに対して有効な表現を学習する方法を提案すること

1 つ目の目的はディープネットによる表現学習が有効であることを示すことでウェアラブルセンサによる行動認識のための表現学習技術を開発するという本論文全体の目的を動機づけるものであり，2 つ目およびに 3 つ目の目的が表現学習の本タスクに対する有効性を高めるためものもである．

本論文は表現学習の観点から見ればウェアラブルセンサによる行動認識の性質を考慮することで，高精度な認識を行うディープネットの応用範囲を広げる試みとして位置づけられる．一方，行動認識研究の観点から見れば，行動認識を行う上での課題であったラベル付きデータが少ないユーザに対する正確な行動認識，およびにプライバシーへの配慮をいかに行うかという課題に対して，表現学習による特徴量の学習というディープネットが登場する以前は難しかった新たなアプローチを検討する試みとして位置づけられるものである．

ディープネットをはじめとする表現学習のウェアラブルセンサを利用した行動認識に対する有効性は，すでにいくつかの研究で検証が進められている．詳細は3章にて説明するが，例えば，Plötz らは行動認識タスクにおいて表現学習をはじめて利用し，従来の特徴設計パイプラインを制限付きボルツマンマシン（Restricted Boltzmann Machine, RBM) や主成分分析 (Principal Component Analysis, 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 [12]．より近年の研究で，Jing らは畳み込みニューラルネット（Convolutional Neural Networks, CNN) 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検証した [19]．また，Ordóñez や Hammerla らは Long Short Term Memory (LSTM) のように時系列性を考慮したディープネットが良い精度を発揮することを確認している [20, 21]．これらのように，ディープネットをはじめとする表現学習はウェアラブルセンサを利用した行動認識において親和性が高く，既存の特徴量と比較して認識精度が高まることが報告されている．

一方で，深層学習の有用性が本分野で認識され始めたのはこの5年余であり，本論文で取り上げるようなウェアラブルセンサの性質を考慮した検討や手法の提案が十分とは言えない．本論文は既存研究の知見を土台に，表現学習がいかにウェアラブルセンサを利用した行動の計測の実用性改善に寄与しうるのか，またその利用可能性を高めるにはどのような技術が必要であるのか，それはどのように実現できるのかについて論じる．また，本論文はウェアラブルセンサによる行動認識を実用的に利用する上での従来からの問題点を表現学習の観点から捉え直し，それぞれの問題に適した手法を提案および検証することで，ウェアラブルセンサによる行動認識に対して深層学習が適用できることを示したものであり，ウェアラブルセンサから得られるパーソナルなデータから実世界の行動を正確に認識することを基礎としたヘルスケア等のシステム発展に貢献するものである．

1.3 本論文の構成

本論文の残りの構成は次の通りである．まず2章では，ウェアラブルセンサを利用した行動認識に関する既存研究をまとめる．ウェアラブルセンサを利用した行動認識をその他の行動認識手法と対比して説明し，ウェアラブルセンサによる行動認識の特徴を明確化する．また，本論文が取り扱うプライバシーへの配慮およびにラベルデータが少ないユーザに対する認識精度をいかに高めるかという問題に関係する既存研究の知見を整理する．3章では，表現学習およびに以降の章で利用する深層学習技術についての基礎知識をまとめる．そのうえで表現学習とウェアラブルセ

ンサによる行動認識に関する既存研究を2章で述べた行動認識研究との対応を整理することで，本論文の学術的な意義を明確化する．

4章では，ウェアラブルセンサによる行動認識の問題に対して既存の表現学習手法であるディープネットを適用し，表現学習により行動認識精度が高まることを確認する．特に，学習データが全く得られないユーザに対する認識精度を評価するユーザ独立行動認識の設定でディープネットと経験的に利用されてきた特徴量を比較することにより，ラベルデータを取得困難なユーザに対する汎用性の観点で評価する．ユーザに関する汎用性は，表現の観点からはユーザの歩行特性の違いに影響されない行動特徴量を利用することで高めることができる．検証にあたり，慣性センサを利用した4つのデータセット合計110万サンプル以上の教師付きデータに対して最大24層のニューラルネットを適用し，その有効性と問題点を実験的に検証した．結果として，ディープネットはユーザの違いに頑健に行動を認識するというの観点で従来利用されてきた経験的な特徴量より優れていることを確認した．一方でディープネットの課題として，単純に深い構造を利用することによる認識精度の改善には限界があること，また獲得された特徴量を分析すると単純なディープネットの適用で得られる表現は必ずしもユーザごとの違いを吸収した特徴量とはならないなどの知見を得た．

5章では，ディープネットを拡張したユーザ敵対的ニューラルネットを提案する．歩行特性などのユーザの違いに影響されないような表現を学習することは，4章で述べるようにユーザ汎用な認識を行うために重要となりうることに加えて，センサにより含まれるユーザあるいはその背後にある属性情報を極力取得しないようにするという意味でプライバシー問題に対しても重要である．本章では，学習時に表現RからユーザUを分類しようとするような敵対的ユーザ分類器を考慮した学習を行うことで，表現Rが持つ認識したい行動に関する情報を最大化しながらユーザに関する情報を最小化するようにディープネットを拡張した手法の提案を行う．3つの行動分類タスクにおいて提案手法と通常のディープネットの認識精度およびにユーザ独立性の比較を行い，提案手法が既存のディープネットと比較してユーザの独立性が高くプライバシーに配慮した表現が獲得できることを定量的・定性的双方の観点から確認する．

6章では，データが少ないユーザに対する行動認識精度を高めるための工夫としてディープネットをユーザ適応する手法の提案を行う．ユーザ適応とは，4章や5章で取り扱ったようにユーザ汎用なモデルを構築することで学習データが少ないユーザに対する認識精度を高めることを目指す代わりに，各ユーザごとに得られる少数

のデータを利用して学習済みモデルをうまく修正することでシステム全体として多くのユーザに対する認識精度を高めることを目指す技術である．このユーザ適応は従来的な特徴設計を利用した手法の枠組みの中で有効性が広く認識されているものの，ユーザ適応する対象をディープネットのように大量のパラメタを持つモデルとする場合には，大量のパラメタを限られたデータでどう適切に修正するかが問題となる．提案手法である半教師あり蒸留とは，学習済みモデルから得られる情報とラベル無しデータを有効活用することで効率的なユーザ適応を行う手法である．既存の 8 つのユーザ適応手法と提案手法を認識精度の観点で比較し，提案手法はラベル無しデータを活用することで過学習を抑制しながら適切なユーザ適応が行えること，ラベルが全く無いユーザに対しても 3%程度の認識精度の改善が見込めること，また既存の手法との組み合わせにより更なる認識精度の改善が見込めることなどを確認した．

7 章では，4 章から 6 章までの実験結果をふまえて，ウェアラブルセンサを利用した行動認識の実世界での利用可能性向上に対する本論文の貢献，および認識精度とプライバシーの観点に対する本論文の限界およびに将来課題をまとめる．最後に 8 章で本論文をまとめる．

また，これらに加えて付録 A では，ウェアラブルセンサを利用した行動認識を応用したシステムとして我々が開発する車いすセンシングを利用した路面アクセシビリティ推定システムについて説明する．ウェアラブルセンサによる行動認識を行うことにより実現されうるシステムの一例を示すことで，4 章から 6 章で取り扱ったウェアラブルセンサからいかに適切に情報を取り出すかという問題に取り組むことでどのようなアプリケーションが実現されるかについての参考としていただきたい．

第2章　ウェアラブルセンサを利用した行動認識に関する既存研究

本論文で扱う行動認識は，ヘルスケアやコンテキストアウェアシステムといったシステムの重要な技術要素として認識されており，ウェアラブルセンサが普及する以前からさまざまな研究がなされてきた．本章ではまず第1に，行動認識を利用するセンサの観点から3つの種類に類型化し，ウェアラブルセンサを利用した行動認識の位置づけを示すとともにその主要な応用範囲について説明する．その他の行動認識技術との対比により，本論文が扱うウェアラブルセンサによる行動認識の範囲およびにその性質を説明する．

第2に，ウェアラブルセンサを利用した行動認識モデルを構築する上での基本的な流れについて説明する．第3に，ウェアラブルセンサを利用した行動認識モデルを評価する上で重要となる認識精度及びにプライバシーへの配慮の観点に関しする既存研究について説明し，本論文で取り扱うデータが少数なユーザを含む不特定ユーザに対する認識精度とプライバシーへの配慮の意義およびに既存研究における知見を整理する．

2.1　行動認識の類型化とウェアラブルセンサの位置づけ

行動認識は1990年代終わりごろに始まった比較的若い研究分野であるものの [22]，その応用範囲の広さからさまざまな研究者が行動認識に関する調査論文を発表している [23, 24, 25]．一方，スマートフォンの登場やセンサの小型化などセンシング環境の変化に伴い行動認識にまつわる用語は変化したり，類似した概念を異なる言葉で指している場合があったりと，調査論文内でも完全な一致は得ていない．行動認識にまつわる用語を整理した上で本論文でのウェアラブルセンサという単語が含む範囲を明確にした上でウェアラブルセンサを利用した行動認識とその他の行動認識の違いを説明し，以降の議論を円滑にすることが本節の狙いである．

まず本論文では,「身につけて持ち歩くことのできるセンサあるいはセンシング機

能を有するもの」をウェアラブルセンサと呼ぶ．これは次のようなものを含む言葉である．

- Jowbone 社が販売している Jowbone Up や東芝が販売している Silmee などのリストバンド型のセンサ
- iPhone や Android 携帯などのスマートフォン
- Apple が発売している Apple Watch や Sony が発売している Wena Watch などのスマートウォッチ
- センサを埋め込んだ衣服 (スマートクローズ) [26]

このような身につけて持ち歩くセンサを利用して持ち歩くユーザの行動を認識することを目的にするのがウェアラブル行動認識である．文献によっては装着型センサ (body-worn sensor) [25] という呼び方がなされる場合もあるが基本的に同じ概念である．

類似する概念として，モバイルセンサを利用した行動認識 [23] や，スマートフォンを利用した行動認識 [27, 28] などがある．これらの行動認識技術はそれぞれモバイルやスマートフォンというある程度の計算能力を持つ機器による計測を前提とするのに対して，ウェアラブルセンサはその処理能力を前提とはしないため厳密に言えば異なる概念である．ただし，モバイルデバイスもスマートフォンも「身について持ち歩くことのできるセンシング機能を有するもの」であるから，本論文はモバイルやスマートフォンなどのセンサを含むより上位の概念としてウェアラブルセンサという言葉を利用する．

ウェアラブルセンサを利用した行動認識と同様に人間の行動を認識することを目的とする技術に，動画ベース行動認識 (video-based activity recognitiion) がある [29]．動画ベース行動認識は，監視カメラなどにより計測した動画像から写っているユーザの行動を認識する技術であり，主に監視やロボットの学習，不審な行動の認識など複雑かつ重要度の高い認識によく利用される．動画ベース行動認識がプライバシーに問題の少ない公共の場など限られた場所での認識を主なアプリケーション領域とするのに対して，ウェアラブルセンサを利用した行動認識はカメラを置けないようなよりプライベートな環境における行動を計測する技術として位置づけられる．

もう 1 つの行動認識技術として，環境埋め込み型センサ (object-embeded sensor) による行動認識がある．環境埋め込み型センサとは，コップや包丁，ドアなどにセンサ機能をもたせることにより，ある環境でどのような行動がなされたかを環境側

から計測することを基本的な目的とする技術である．また，通常環境埋め込みセンサはそれ自体がセンサとしてわからないように環境に溶け込む形であることが多いため，アンビエントセンサ (ambient sensor) と呼ばれる場合もある．環境埋め込みセンサは環境側にセンサを設置するという観点で動画による行動認識と類似しているが，プライバシーへの配慮の観点から利用されるセンサは動画と比べるとより低レベルなものであることが多い．また，プライバシーに配慮したデータ取得方法でありプライベートな空間での利用を想定しているという点において環境埋め込みセンサはウェアラブルセンサと類似しているものの，環境埋め込みセンサを利用した行動認識が特定の環境における行動を計測することに主眼を置くのに対して，ウェアラブルセンサを利用した行動認識は特定のユーザの行動を連続的に記録することに主眼を置く技術であるという点で異なっている．

以上をまとめると，本論文ではウェアラブルセンサを「身につけて持ち歩くことのできるセンサあるいはセンシング機能を有するもの」として定義する．また，ウェアラブルセンサを利用した行動認識はその他の行動認識技術と比較して次のような特徴がある．

ウェアラブルセンサを利用した行動認識の特徴

1. ユーザと行動（観測データ）が紐づくデータを対象とする
2. プライベートな空間における行動の記録を目的とする

また，このようなデータを対象とする性質上，認識精度を高めるだけでなく，プライバシーへの配慮が必要となる点がウェアラブルセンサによる行動認識の技術的な特徴である．

やや区別が曖昧な例として，ウェアラブルセンサと環境埋込み型センサを併用している場合がある．例えば，Pansiot らはヘルスケアモニタリングシステムを作る上でウェアラブルセンサで認識できない行動を識別することを目的に環境に埋め込んだセンサのデータと組み合わせることを提案しその有効性を検証している [30]．ただし，これらのようにウェアラブルにおける行動認識の精度を向上させる目的で環境埋込み型センサを利用する場合上記に述べたようなウェアラブルセンサを利用した行動認識の特徴を基本的には含む．そのため，本論文ではウェアラブルと環境埋込み型センサを併用する場合はその目的がウェアラブルセンサによる認識精度の補完である場合にはウェアラブルセンサによる行動認識の対象として扱う．

ウェアラブルセンサを利用した行動認識により実現されるアプリケーションの代表的なものには次のようなものがある．

活動量計

すでに実用的に利用されているアプリケーションの事例としては活動量計がある．Joubone Up などのリストバンド型のセンサや Apple Watch や Wena Watch などのスマートウォッチ，あるいは iPhone などのスマートフォンなどの多くで活動量を計測・可視化する機能を有している．また，計測した活動量を単に可視化するのみならず，その活動量の履歴から運動に関するアドバイスを行うようなシステムも存在する．

ヘルスケア

ヘルスケアは，ウェアラブルセンサを利用した行動認識の最も主要な応用先の 1 つでありさまざまなアプリケーションが提案されている．Sung らは，パーキンソン病患者のモニタリングやてんかん性発作 (epileptic seizures) の検出を試みている [2]．Steele らは，慢性閉塞性肺疾患 (chronic pulmonary disease) 患者の行動をセンサから認識することで，リハビリテーションの補助や研究による病理理解を補助することを提案している [3]．あるいは，肥満患者の計測した行動履歴を元にライフスタイルや行動の改善など治療の補助に利用するシステムも提案されている [4, 5]．身体的なデータのみならず，Ertin らは，音声センサや加速度センサ，心拍センサなどを利用してユーザのモニタリングを行うことで，そのストレス状態を 90%以上の精度で推定可能であることを示している代表的な調査論文として，[31] がある．

デッドレコニング

近年注目を集めている応用先の 1 つに，デッドレコニング (dead reckoning) がある [32]．デッドレコニングとは，加速度センサやジャイロセンサ，気圧センサなどを利用してユーザの動きから相対的な移動距離を推定することでユーザの位置を推定する技術であり，特に屋内など GPS が利用不可能な場所における位置測位を実現する技術として期待されている．デッドレコニングはユーザの行動から相対的な位置を推定するものであり絶対的な座標をそれだけから推定することはできないが，Wi-Fi などのその他の位置測位手法を組み合わせることで座標情報を推定することができる．2015 年に行われた UbiComp (International Conference on Ubiquitous Computing) では，

デッドリコニングに関するチャレンジ (Ubicomp/ISWC 2015 PDR Challenge[1]) が行われ，100 名以上の参加者が参加している．

実世界マイニング

実世界マイニングとは，通常のセンサでは計測できない現実世界の状態を何らかの形で計測し情報を取り出す仕組みのことである．Web マイニングが Web を観測し情報を取り出すことを指すのに対して，実世界マイニングは実世界を観測し情報を取り出す研究である．行動認識によって得られる情報は，あるユーザがどこで何をしていたかという情報であり，実世界マイニングの重要な情報の 1 つとして認識されている．代表的なアプリケーションとしては，ユーザの移動先推定に基づく混雑予想システムや，病院における看護士の行動のセンシングに基づく最適行動パターンの提示システムなどがある．

このように，ウェアラブルセンサを利用した行動認識に基づくアプリケーションは多様であり，種々のアプリケーションにおいて認識した情報をどう活用するかについてはアプリケーションによる個別性が高く一般的な技術開発は難しい．一方，いずれの応用領域においても共通する一般的な流れと，ウェアラブルセンサを利用して実空間における人間行動を計測するという目的から生じる共通の難しさが存在している．以降，まず 2.2 節ではウェアラブルセンサを利用した行動認識の一般的な流れを整理する．この一般的な流れをふまえて，2.3 節以降ではウェアラブルセンサを利用した行動認識モデルを評価する軸として，本論文が扱う認識精度およびにプライバシーへの配慮の観点に関連する既存研究について述べる．

2.2 ウェアラブルを利用した行動認識モデルの構築方法

ウェアラブルセンサによる行動認識は，パターン認識問題 [33] や記号推論の問題として扱われる場合もあるが，最も一般的にはある時刻 t におけるセンサデータを入力 X にとりその時刻 t における行動 y を予想する教師あり学習問題としてとして定式化される．文献によりやや用語の用いられ方は異なっているものの，行動認識モデルの学習は一般に次に示すような手順で行われる（図 2.1）

[1] http://pdr.hasc.jp/

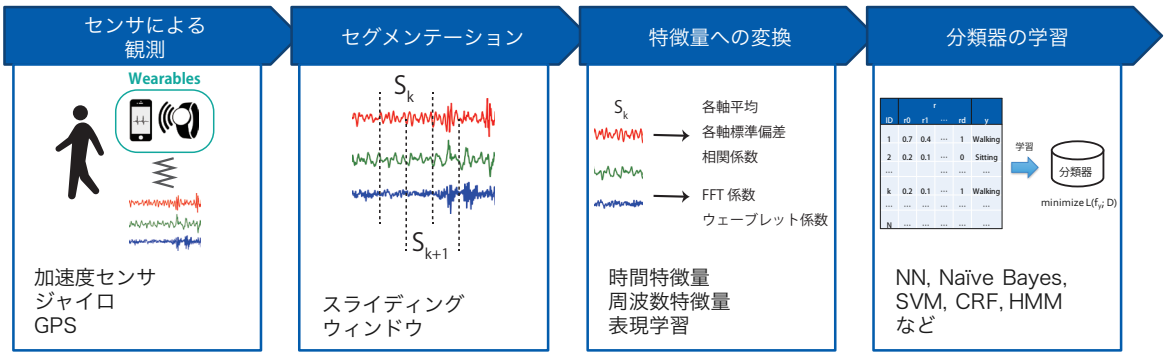


図 2.1: ウェアラブルセンサを利用した行動認識モデル構築の一般的な流れ.

1. センサによる行動の観測：ウェアラブルセンサを利用してある時刻におけるユーザの行動を計測した観測値 X を得る.
2. セグメンテーション：観測信号を機械学習で扱いやすいようにセグメント S_k に分割する.
3. 特徴量への変換：セグメント S_k を何らかの特徴ベクトル r_k に変換する.
4. 分類器の学習:獲得された特徴ベクトルと行動ラベルのペア $\{r_k, y_k\}(k = 1, 2, \cdots)$ を利用して，r_k から y_k を予想するためのモデルを学習する.

以降それぞれの観点について説明する.

2.2.1 センサによる行動の観測

行動認識の定義上自明であるが，まず最初のステップはセンサによる行動の観測である．以降，観測されたセンサデータの系列を

$$\mathbf{X} = [\cdots, \mathbf{x_t}, \mathbf{x_{t+1}}, \cdots]$$

と表記する．ただし，$\mathbf{x_t}$ は時刻 t おける複数のセンサによる観測値である．例えば，3 軸加速度センサを利用する場合には $\mathbf{x_t}$ の次元は 3 であり，3 軸加速度センサに加えて 3 軸ジャイロセンサを利用する場合にはその次元数は 6 である．実際のところ，センサによるサンプリングレートの違いや複数のセンサを利用する際の同期をどのように行うかという問題があるため観測時刻 t や観測数はセンサによりばらつく可能性もあるが，以降の議論に影響しないため補完などの方法により各センサの観測時刻 t およびに観測数は同一になっていると考える.

一般に，利用するセンサにより観測可能な情報は異なるので，アプリケーション目的に応じたセンサの選択が必要となる．ウェアラブルセンサによる行動認識で利用されるセンサには大別すると次のようなものがある．

加速度センサ

ウェアラブルセンサを利用した行動認識を行う場合に最も利用されるのが加速度センサである．特に，歩く，走る，座る，階段を登るなどの基本行動をはじめとした物理的なユーザの行動を推定する場合によく利用される．加速度を利用した行動認識を行っている研究としては，例えば [34, 35, 36, 37] などがある．

ジャイロセンサ

ジャイロセンサとは，別名角加速度センサとも呼ばれるセンサであり，物体の角度や各加速度を計測するのに適しているセンサである．行動認識の分野では特にユーザの姿勢推定 (pose estimation) によく利用される [38]．また，加速度センサとの組み合わせて利用されることが多い [39, 38, 40]．

GPS

本論文では実例としては扱わないが，GPS (Global Positioning System) を利用した行動の認識もよく行われている．特に，行動認識がジェスチャや身体の動きなどの身体の行動を検出することを主な目的とするのに対して，GPS による行動認識では移動モードの推定などより高レベルの認識が行われることが多い．

生体センサ

上記のセンサに比べると最近になって利用が始まったセンサの 1 つが生体センサである．生体センサは多種多様であり，例えば血圧センサ，心拍センサ，EEG，ECG，呼気センサなどがある．例えば，Sung らは，兵士の体温をモニタリングし低体温を検出するシステムを提案している [41]．Hossain らは，呼気センサを利用して喫煙行動を検出しユーザに対するフィードバックにより喫煙を抑制するシステムを提案している [42]．

2.2.2 セグメンテーション

セグメンテーションとは，生のセンサデータを理想的には各セグメントが異なる行動の情報を含むように分割するプロセスのことを指す．セグメンテーションは後に行う分類タスクにおいて有用であるだけでなく，センサに変化が少ない場合など認識する必要が無いと思われる場合に分類を行わないなどバッテリー消費を抑える目的でも利用される．セグメンテーションでは，何らかの前処理を行われた X から系列をサブ時系列 $\mathbf{S_k}$ を切り出す．

$$\mathbf{S_k} = [\mathbf{x_t}, \cdots, \mathbf{x_{t+L-1}}]$$

ここで，L は切り出す系列の長さであり，$\mathbf{S_k}$ は理想的には何らかの行動と対応する時刻部分を $\mathbf{X}$ から切り取ったものである．

ただし，実際にはセグメントは行動やコンテキストを認識するための前段階として行うものであるから，この段階で行動をセグメントすることは一般には難しい．そのため，前から少しずつずらしてデータを切り出すスライディングウィンドウ方式 (sliding window procedure) が最もよく利用される [25]．スライディングウィンドウ方式とは，長い系列を前方から順に固定長の窓を徐々に時刻方向に滑らしていくことにより複数のサブ時系列を切り出す方法である．このようにして得られた各 S_k に対して，それぞれ次ステップの特徴抽出を行う．

2.2.3 特徴量への変換

特徴量への変換ステップでは，切り出されたされた系列データ $\mathbf{S_k}$ を何らかの関数により認識したい行動を識別しやすいようなデータ表現に変換する．一般に，行動認識に限らず機械学習を利用する多くのタスクにおいて特徴量の設計はその認識精度に大きな影響を与えるプロセスであることが知られている．例えば，複数の特徴量で定義される特徴空間において，同じ行動は近くに集まり異なる行動は遠く離れるような特徴量を定義することができればその分類を行うことは簡単である．一方で，特徴空間で複数の行動が同じような特徴量を取る場合，基本的にはどのような分類器を利用するかにかかわらずその分類は難しい．本論文が取り扱う表現学習は，従来的な機械学習が特徴量への写像関数 f を人手で設計するのに対して，何らかの目的関数を最適化するような関数 f をデータから発見する方法である．ウェアラブルセンサを利用した行動認識に対して表現学習を適用した事例はいくつか存在して

いるものの，殆どの研究は人手による特徴設計を利用している．ここでは人手により経験的に設計されたきた特徴量の代表的な例に言及する．

特徴量への変換の工程は定式的には，

$$\mathbf{r_k} = \mathbf{F}(\mathbf{S_k})$$

のように各セグメント S_k に対して何らかの関数による変換を加えることである．ここで F は何らかの特徴抽出器，あるいは特徴抽出器の組み合わせである．行動認識においてよく利用される特徴量には例えば次のようなものがある．

時間領域統計量

最もよく利用されるのは各信号系列の各セグメントの各入力チャネル（例えばX軸加速度，Y軸加速度など）ごとの時間領域における統計量である．例えば，平均，標準偏差，最大値，最小値などが用いられる．統計量は計算コストおよびに必要なメモリも少なく，また加速度センサやジャイロスコープなどさまざまなセンサで利用される．特に，運動強度により識別が可能な場合には時間領域統計量がよく利用される．

相関係数

相関係数は2つのベクトルの類似度を表す指標であり，$corr_{x_1,x_2} = \frac{cov(x_1,x_2)}{\sigma_1,\sigma_2}$ により計算される．ただし，x_1 およびに x_2 はセンサの系列であり，$cov(x_1, x_2)$ は x_1 と x_2 の分散共分散行列，σ_1 およびに σ_2 はそれぞれ x_1 と x_2 の標準偏差である．

相互相関係数

複数のセンサのデータ間の関係を捉える特徴量として，相互相関が利用される．相互相関は，2つの信号の類似度を図るためによく利用される指標である．相互相関は，次のように2つの信号のドット積の最大値として得られる．

$$CrossCorrelation_{x,y} = \max_{d=1}^{n-1}(\frac{1}{n}\sum_{i=1}^{n} x_i y_{i-d})$$

典型的には，観測信号が k 個のセンサであった場合そのペアワイズでの相関が利用される．例えば，Veltink らは1軸加速度センサを利用して相互相関係数を利用し

たテンプレートマッチングを行うことで立つ，座る，寝そべる，歩く，階段をのぼる，降りるなどの行動を識別する方法を提案している [43].

Empirical Cumulative Distribution Function (ECDF)

比較的近年提案された方法として，ECDF と呼ばれる方法がある．ECDF とは Empirical Cumulative Distribution Function の略であり，日本語に訳すと経験累積分布関数となる．この関数では，あるセグメント内の累積分布を得られた代表点により近似的に表現する．例えば，ベクトル $\mathbf{z} = [\mathbf{z_1}, \mathbf{z_2}, \cdots, \mathbf{z_n}]$ が与えられた場合，ECDF は次の値で計算される要素のベクトルにより表される．

$$ECDF_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\mathbf{z_i} \leq \mathbf{t}}$$

ただし，$\mathbf{1_A}$ は事象 A が満たされている場合に 1 を満たされない場合に 0 を返すインデックス関数である．ECDF では，複数の t において上記の値を計算したベクトルを作成する．

ECDF は，Plötz ら [12] により初めて提案され，Hammerla [11] らにより複数のデータセットでの有効性が検証されている．詳細は，[11] を参照されたい．

周波数領域特徴量

ここまでは時間領域においてそのまま特徴量を計算する特徴量であったが，時間領域でそのまま特徴量を計算するのではなく高速フーリエ変換 (Fast Fourier Transformation, FFT) などを利用して一旦周波数領域に変換した上で特徴量を計算する場合もある．例えば歩き動作や走り動作などの一部の行動は繰り返し動作の場合には特定の周波数成分が現れやすくなると考えられ特に有効である．周波数成分への変換に最もよく利用されるのは高速フーリエ変換であるが，種々の母関数を利用したウェーブレット変換も利用される [44, 45].

周波数領域における各周波数成分の強さ (周波数計数) が直接利用される [46] こともあるが，直流成分の強さ [34]，エネルギー [47]，エントロピー [34] などが利用されることもある．

2.2.4 分類器の学習

最後の工程が分類器の学習である．ここでは，ここまでの手続きにより計測された生データから算出された特徴量 r と行動のラベル y を利用して，特徴量とラベルの対応関係を学習する．対応関係がどのように表わされるか (関数 $y = f(r)$ として表わされるか，事後確率 $P(y|r)$ として表わされるか，あるいは $P(r)$ もモデル化するか)，最適に利用される目的関数などは選択する分類モデルにより異なるものの，基本的には通常の教師あり学習の問題として定式化できる．そのため，基本的には既存の機械学習アルゴリズムの適用が主であり新しいアルゴリズムの開発が行われることは多くない．よく利用される認識アルゴリズムとしては，決定木，k 近傍法，ニューラルネット，SVM，隠れマルコフモデル，Conditional Random Field などがあり [25, 23]，各アプリケーション開発時には種々のアルゴリズムを比較して選択することが行われている．

分類モデルの学習にまつわる問題として，ラベルデータをいかに入手するかという問題がある．特に訓練データセットを実験環境で作成する場合，ラベルデータの作成には実験時に作成した動画が利用されることが一般的である．画像認識や音声認識などの領域と比較して，ウェアラブルセンサを利用した行動認識におけるラベルデータの確保は難しい．この難しさは，次のような理由によるものである．

- 解釈困難性．画像であれば視覚的に，音声であれば聴覚的に人間が判断可能である．そのため，画像や音声は観測信号のみから事後にラベル付けが比較的容易である．一方，ウェアラブルセンサを利用した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ープは，その信号のみから事後にラベル付けを行うことは困難である．

- プライバシー．実環境におけるラベルの取得に動画を利用することは，ウェアラブルセンサによる行動認識の持つプライバシーへの配慮という性質と相反するものであり通常は難しい．

- 正例観測の難しさ．例えば，転倒検出を行いたい場合に転倒に対する教師データを取得するには実際に転倒してもらう必要がある．擬似的なデータであれば例えばマットレスを用意して安全な環境で転倒してもらう等のことが考えられるが，より自然なデータを記録しようとすると転倒の観測回数は限られる．そのため，実環境で取得が困難なデータが存在する．

これらの理由から，画像認識や音声認識などと比較してラベル付きデータの確保は難しく，それゆえいかに限られたデータから効率的に学習するかが重要となる．

2.3 ウェアラブルを利用した行動認識モデルの評価

2.3.1 ユーザ独立認識精度とユーザ依存認識精度

ウェアラウルセンサによる行動認識モデルの目的は行動を高精度に識別することであり，どの程度正確に行動を認識できるのかが第一の評価指標になる．ウェアラブルセンサによる行動認識モデルを評価する場合，どのようなユーザを認識対象とするかによってユーザ独立認識と呼ばれる問題設定とユーザ依存認識と呼ばれる問題設定に分類される．ユーザ独立認識は，モデルの学習にデータが利用されていないユーザを含む不特定ユーザの行動を認識することを，ユーザ依存認識はモデルの学習にデータが利用された特定のユーザ群の行動を認識することを目指すものである．例えば，モデルの学習に利用可能なユーザとして A, B, C がいる場合を考える．この時，ユーザ依存認識は A, B, C のみを認識対象として考えるような問題設定である．ただし，これは全データが観測可能であるということではなく，A, B, C のユーザに対して得られる新たなデータに対する認識精度を検証する設定であることに注意されたい．一方ユーザ独立認識は，A, B, C だけではなく学習時には存在しなかった，あるいは学習データが得られないようなユーザ D, E, F, $\cdots$ も認識対象として考慮するような問題設定である．

行動認識においてこれらが明確に区別される理由は，主に適用可能範囲の違いと難しさの違いにある．まず，認識精度の観点ではユーザ独立認識はユーザ依存認識とくらべて著しく低いことがいくつかの既存研究により報告されている．以下にいくつかの研究事例を列挙する

Bao らは 20 名の行動データセットを利用した認識実験を行い，ユーザ独立認識とユーザ依存認識の認識率の観点から見た比較を行っている [34]．認識対象としては，歩く，座る，立つ，テレビを見る，エレベータに乗るなど合計 20 個の行動を対象にしている．結果として，同程度のデータ量を利用する場合にはユーザ依存認識の方が認識精度が高いことを報告している．

Weiss らは，59 名の歩く，ジョギングする，階段をのぼる，座る，立つ，寝そべるの 6 つの行動合計 9,291 個のサンプルを利用した実験によりユーザ独立認識とユーザ依存認識の精度を比較している [48]．分類器としては，WEKA に搭載されているア

ルゴリズムを対象としており，J48 やランダムフォレストなどの決定木，ニューラルネット，ナイーブベイズ，ロジスティック回帰など合計 8 種類の手法で検証している．結果として，ユーザ依存認識においてはニューラルネットが最もよく 98.7%（ユーザ依存認識では 67.8%） である一方でユーザ独立認識においてはランダムフォレストが最もよく 75.9%（ユーザ依存認識では 98.4%） であったことを報告している．

類似した研究として，Lane らは 50 名の行動データセットを 20 代 30 代のデータと 65 歳以上のデータセットに分けて分析を行っている [37]．具体的には，20 代 30 代のデータにより訓練した認識モデルの精度を同じ 20 代 30 代のユーザのデータの識別に利用する場合と 65 歳以上の行動を識別する場合とで認識精度を比較している．結果，20 代 30 代に対しては 80%から 90%達成するモデルが 65 歳以上のユーザに対しては 65%程度まで認識精度が低下することを報告している．

これらの研究はいずれも，ウェアラブルセンサにより観測される値はユーザに関する何らかのバイアスを持ったものであることを意味している．すなわち，同じ行動をする場合でもその行動に対応した観測値がどのようになるかはユーザにより異なる．

そのためユーザ依存認識は認識精度の観点からすると優位であるものの，ユーザ依存モデルを利用した認識を行うためには認識したいユーザごとにラベルデータを確保する必要があるという問題がある．特に多くのユーザを認識対象とするアプリケーションではコストやプライバシーの観点から現実的にはそれぞれのユーザに対してラベルデータを確保することは難しい場合が多い．そのため，多くのユーザの行動を認識するようなアプリケーションを想定する場合にユーザ依存認識は適用困難である．そのため，ユーザ独立認識の精度を高める仕組みが必要とされている．

また，認識モデルをどのように構築するかにより行動認識システムは次の 3 つに大別することができる．第 1 に個人モデル (personal model) とは，特定の 1 人のユーザで訓練されたモデルのことである．例えば，ユーザ A の行動を識別したい場合にはユーザ A について集めたデータのみを利用して学習を行うという方法である．個人モデルは各個人の特徴を捉えたような識別を行えば良いためユーザに関する汎化性能が求められない分認識精度は比較的高くなる一方で，認識対象とするユーザそれぞれにラベル付きデータが一定量必要となる．

第 2 の，非個人モデル (impersonal model) とはなるべく多くのユーザのデータを利用して学習されたモデルのことである．例えば，ユーザ A，B，C の学習データがある場合に個人モデルのアプローチを採用する場合にはユーザ A，B，C それぞれのデータを作って 3 つの異なるモデルを構築するのに対して，非個人モデルのア

プローチを採用する場合にはユーザ A，B，C 全員分のデータを利用して 1 つのモデルを構築する．なお，ユーザ依存認識の設定で利用される場合ハイブリッドモデル (hybrid model) と呼ばれることもある．これまでに述べたユーザ独立認識とユーザ依存認識に関する比較実験を始めとして，ユーザ独立認識を行う場合には非個人モデルが最もよく利用されている．実際のところ，非個人モデルが認識精度の学習時に含まれないユーザに対する認識精度の観点で優れているかどうかは自明ではないものの，ユーザ独立認識の場合に非個人モデルが個人モデルより好まれるのには次のような理由がある．

第 1 に非個人モデルは個人モデルより大規模なデータにより学習することができる．例えば，10 人分の学習データがそれぞれ 1 時間ずつある場合，個人モデルはそれぞれ 1 時間のデータで訓練したモデルを 10 個作成するのに対して，非個人モデルでは 10 時間のデータを利用して訓練することができる．あるいは，同じ時間に絞る場合はデータの取得コストを複数のユーザに分散することができるため効率が良い．第 2 に，特にモデルの表現力が限られている場合，すなわち効率的に学習データを分類するような規則を獲得する必要がある場合，非個人モデルは学習データに含まれる複数のユーザのデータ共通に利用可能な規則を学習する必要がある．あるいは，モデルの表現力が十分にある場合にはいずれのユーザでも認識できるような複雑な規則を学習することができる．

一方，すでに見たように非個人モデルは個人モデルと比較して通常認識精度の観点で著しく劣ることが知られている．この理由はいくつか考えられるが，その理由の 1 つに特徴量のユーザ依存性が考えられる．特徴量がユーザ依存していることが認識精度に与える問題について，学習に利用するユーザが 2 人（赤及びに青）で，認識したいユーザが 1 人（緑），認識対象の行動が 2 クラス（ダイヤモンド図形およびに丸型図形）の場合で例示する (図 2.2)．ここで，点線は擬似的に記した分類超平面を表しており，訓練ユーザに対しては 100%の精度となるようなものである．一方，この平面は緑ユーザのサンプルに重なっており，緑ユーザに対してはあまり良い認識精度を達成しないことがわかる（実際には 70%）．

この例は擬似的な例であるが，Lane らは実データを利用した検証により特徴量のユーザ依存性は実データにおいても存在することが報告されている [37]．具体的には，先ほどと同様のデータを利用した 50 名の行動データセットを利用して，20 代 30 代が歩いている時の特徴量と 65 歳以上の歩いている時の特徴量をプロットして比較し，両者の特徴量がほとんど重ならないことを報告している．特徴量としては先行研究 [49] で利用されている行動認識に有効であることがすでに検証済みのもの

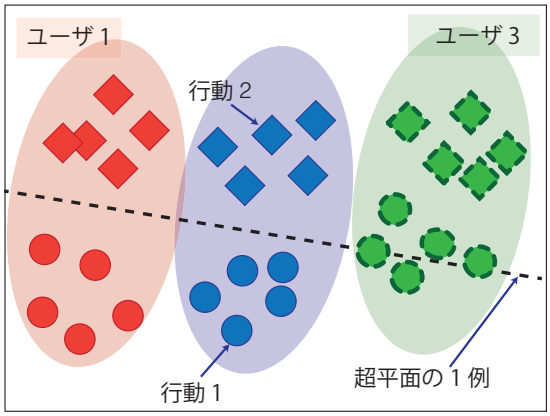


図 2.2: 特徴量のユーザ依存性がもたらす問題．赤，青，緑がそれぞれ異なるユーザを，ダイアモンド型のサンプルが行動 1，丸型のサンプルが行動 2，また点線が赤青ユーザに対して 100%の精度で認識可能な超平面の 1 例を示している．

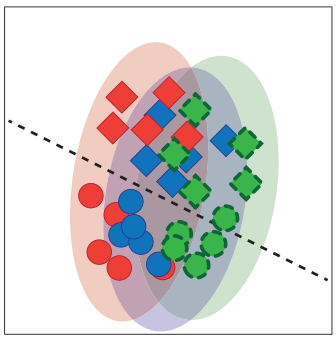


図 2.3: 特徴量のユーザ独立性がもたらす利点の概念図．

を利用し，主成分分析により 2 軸に変換して可視化したと記述されている．詳細は [37] の Figure1 を参照されたい．

逆に，ユーザに依存せずに同じような値を取る特徴量を利用することができれば，この問題は軽減することができる．図 2.3 に，同じデータ例を利用して，ユーザに依存しない特徴量を利用することの利点を説明する．点線は，先ほどの場合と同様に擬似的に記した分類超平面であり，訓練ユーザに対しては先ほどと同様 100%の精度である．また，先ほどは緑ユーザに対して認識精度が大きく劣化していたが，今回の場合は 90%の認識精度となる．このように，ユーザ依存性の低い特徴を利用することで，すでに学習された分類規則がその他のユーザに対しても適用可能になる蓋然性は高く，結果として不特定ユーザに対する認識精度の改善にも繋がると期待できる．

特徴量の観点から非個人モデルを考えると，より多くのデータを利用して学習する非個人モデルは特徴空間中のさまざまな場所を網羅するようなデータを使用して学習を行うことを意味する．実際，特徴量の全空間を学習データが網羅しており，かつ学習データに対して良い認識精度を達成するようなモデルが学習できれば，そのようなモデルは多くのユーザの行動を認識できる可能性は高いため，学習に利用するユーザ数を増やすことは理にかなっている．一方で，このアプローチは 2 つの意味で限界がある．

第 1 に，特徴量空間 R の広さである．特徴量空間 R の次元は，特徴量として利用する数により決まる．そのため，認識に利用する特徴量を増やせば増やすほど特徴空間 R は広くなる．特徴量は増やせばいいというものではないため，実質的には何らかの数で抑えられるものであるものの，特徴量が少なすぎれば高い精度で認識することは難しくなる．特徴量空間 R の広さは，特徴量の次元に従い指数関数的に増加していくため，網羅することは難しい．そのため，特徴空間を網羅するようなデータを集めることは少なくとも非常に難しい．

第 2 に，仮に多くを網羅することができたとしても，幅広く分布した特徴空間を正しく分離するような平面を学習することは難しい．すなわち，ユーザを増やせば増やすほどその入力と出力の対応関係が複雑になる．例えば，Ogawa らは，ユーザ独立モデルの学習に利用するユーザ数を 1 人から 60 人まで増やしてその認識精度の変化を比較し，40 人以上を増やしても認識精度は大きく改善しないことを報告している [50]．そのため，そもそも学習に含まれる全ユーザに対して一定以上の認識精度を持つような非個人モデルを学習させることは少なくとも表現力の小さなモデルでは難しい．

個人モデルと非個人モデルの中間的なモデルとして，個別化モデル (personalized model) がある．ユーザ独立認識の精度を高めることはアプリケーションを提供可能な潜在ユーザを広げることになるため実用上重要であるが，行動の仕方はユーザごとに異なることから単一の非個人モデルにより認識精度を担保することは難しい場合が多い．個別化モデルとは，何らかの方法により取得した特定ユーザのデータ（一般的にはごく少数のラベルづきデータやユーザの属性情報）を利用して，学習済みのモデルを微調整することで特定のユーザに対して特化するように修正されたモデルのことを指す．このように，学習済みのモデルを修正する技術はユーザ適応 (user adaptation) と呼ばれ近年注目されている．言い換えると，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，転移学習の一種として捉えることができ

る．ユーザ適応を行う利点としては，主に次の 2 つである．

まず第 1 に，非個人モデルと比較すると，個人に適合するように修正されたモデルであるため，より正確な認識が期待できる．第 2 に，個人モデルと比較するとユーザごとに用意する必要のあるラベルは少数ですむ．すなわち，学習済みの非個人モデルの知識を利用するため，データが少ないユーザに対しても正確な認識ができることが期待できる．

ユーザ適応に関する具体的な研究として例えば，Foster らは，行動認識の結果に対するユーザのフィードバック情報を利用してサンプルの重みを変更する Incremental KNN を提案している [51]．Parkka らは，あるユーザのデータで訓練された決定木の分類規則 (閾値) を新しいユーザの教師ありデータを利用して修正することでパーソナライゼーションを行う方法を提案している [52]．Zhao らは，教師ありデータを利用しない方法として k 平均法によるクラスタリングを利用して決定木の規則を更新する方法を提案している [53]．Hachiya らは，学習データのあるユーザと教師付きデータのないユーザの違いをドメインの違いであると捉え，ドメイン適応の手法の 1 つである共変量シフト仮定に基づいたサンプルへの重み付けを通じて教師データの無いユーザに対する認識精度を改善する方法を提案している [54]．また，Reiss らはより複雑なモデルに対するパーソナライゼーションをモバイル端末側で行うための方法として，アンサンブル学習した複数のモデルの重み付けのみを新しいユーザのデータで更新する方法を提案している [55]．

2.3.2 プライバシーへの配慮

行動認識モデルを評価する上で認識精度の他に重要な観点としてプライバシーへの配慮がある．歴史的に見ると，プライバシー保護に関する研究の多くはユーザの属性情報に関するデータベースや，あるいは位置や音声などの直接人間が解釈可能なデータに関するものであった．しかし，アルゴリズムの進歩により認識精度が高まるにつれてその他のセンサに関するプライバシーへの配慮の必要性が認識され始めている．行動認識を利用したアプリケーションではユーザの同意に基づくデータの収集・解析・共有が前提となるため，プライバシーに配慮することはアプリケーションの利用者を増やし価値を高めるために重要である．

ウェアラブルセンサによる行動認識は加速度センサなどの生データを観測するセンサを身につけて行動を計測するという性質上，次の 2 つの問題を含みうる．

第 1 に，ユーザと行動が直接紐づくという問題である．これは，ヘルスケアなど

のサービスを行う上では必要であるものの，データベースが漏洩すると特定のユーザがどのような行動をしたかが漏れ無くわかってしまう．そのため，多重仮名などの技術的な対策が必要となる．

また，多重仮名などの対策を行っていた場合でも，データベースに存在するその他の要素からユーザが識別できる可能性がある．データベースのうち特定の要素あるいは複数の要素の組み合わせによりユーザの識別が可能であるとき，そのような組み合わせは一般に擬似 ID と呼ばれる．疑似 ID の最も有名な例として，例えば米国のデータでは（性別，生年月日，住所の郵便番号）の 3 つの情報のみから全国民の 85%が一意に絞りこめると推定されている．

このようなデータベースに入ったテーブル上のデータからのプライバシー漏えいを避けることに対する技術的対策は，プライバシー保護と呼ばれる領域で 2002 年頃から行われてきた．代表的な手法としては，匿名化，差分プライバシー，検索クエリ監査，暗号計算などの技術がある．それぞれ異なる目的をもつ技術であるが，基本的に行動認識問題においても同様の手続きが適用可能であり，行動データに対しても同種の手続きを加えることでプライバシーの保護を行うことができる．

より新しい指摘として，生データの持つ情報量の大きさに起因する情報暴露の可能性の危険性が指摘されるようになった [56, 57, 58] これは，アルゴリズムやさまざまな実証実験により生データからユーザの行動や心理状態，あるいは属性までもが抽出可能であることがわかりはじめたことにより生じた比較的新しい問題である．Raiji らは生データからの情報暴露の危険性がプライバシーの新しい問題であること，またその危険性が一般にユーザにあまり認識されていないことをアンケート調査により検証した [56]．Han らや Chakraborty らは，加速度センサは位置情報の推定やあるいはパスワードの推定に利用可能であるにもかかわらず画像情報や音声情報とは違い保護されておらず，アプリケーション開発者が自由にアクセスできることへの危険性を指摘している [59, 60]．

ユーザのプライバシーに関わりそうな情報の推定事例として，例えば Brdar らや Itay らは，加速度センサの値を含むモバイルデバイスにより計測可能な素性から性別や年齢などの属性情報の推定を試みている [61, 62]．Han らは，位置情報が仮に利用不可能であったとしても加速度から復元した情報のみで 200m 程度の誤差範囲で位置情報の推定が可能であることを報告している [59]．Owusu らは，加速度センサにより抽出した特徴を利用することでスマートフォン上の画面の触れられているかを推定可能であることを確認し，パスワード流出などの危険性に言及している [63]．Plarre らは，EEG や RIP (respiratory inductive plethysmograph) により計測したセ

ンサの観測値とユーザがレポートするストレス度合いに相関がある (0.71) ことを報告している．また，ヘルスケア領域に関するアプリケーション事例の項目でも述べたように，さまざまな研究で加速度センサから健康状態を示しうる行動が推定可能であることを報告している．

この問題に対する技術的対策として，Chakraborty らはセンサから推定されうる潜在的な情報をユーザに提示することで，ユーザがプライバシー情報の漏洩リスクを管理しやすくなるような仕組みを提案している [60]．Choi らは，ユーザのデータを中央集権サーバにすべて集めるような方法に対して，ユーザのデータを分散させた上でユーザがその嗜好に応じてどの情報を渡すかを動的に決定できるようにする方法を提案している [64]．Choi らの方法は，データの所有権がユーザあることを担保しようとするという意味で，カプアキンらが提唱するプライバシーデザイン (Privacy by Design) [65] の思想に近く興味深い．情報のコントロール権はプライバシーの基本的な考え方の 1 つであり，収集されたデータのコントロール権をどのように担保するのかというのは技術的には重要な問題である．

一方，この問題の難しい点の 1 つは，プライバシーに対する保護の強さとデータの価値が究極的には相反する関係になることである．例えば，Chakraborty らの方法により特定のセンサの利用を停止することは可能であるが，その場合特定のサービスが受けられないことを意味する．これは Choi らの方法も同様である．

この問題への解決策として Raghavan らは，ユーザの状況に応じたプライバシーコントロールを提案している [66]．これは例えば，ユーザが病院付近にいる時にはその情報を欠落させるといった対策であり，ユーザがどのようなコンテキストにおいてどのようなセンサの情報をアプリケーションに公開するかを細かく指定した上で，コンテキストの推定を行うことで実現できる．この方法では確かにプライバシーと情報の価値のバランスをコンテキストという条件付けの中で調整できるため従来法より柔軟ではあるものの，基本的に問題の本質は変化していない．

もう 1 つの方法として，Chakraborty らは特徴量を難読化 (obfuscation) することを提案している [67]．具体的には，特徴量をすべて使うのではなく本当に予測したい（あるいは予測することをユーザが同意した）情報の予測は可能だが機微情報の予測は難しくなるような特徴量を選択する方法を提案している．このような特徴量を設計することができれば，単にセンサを利用するかしないかという粒度ではなくある特徴を利用するかしないかというより細かな粒度でプライバシー問題を考えることができる．

一方この方法ではそもそもの特徴量のセットとして開発者が設計した情報しか利

用できないため認識精度に課題が残る可能性がある．また，表現学習の文脈では表現は固定されたものではなく，学習されるものであり可変であるため直接特徴選択の枠組みで議論することは難しい．本論文では特徴量の難読化を表現学習の枠組みで実現する方法として，特定の因子（行動など）については予測精度を持ちながら機微情報については予測精度を持たないような表現を学習する方法として定式化しその有効性を検証する．

第3章　ウェアラブルセンサを利用した行動認識のための表現学習

2 章では，本論文が扱うウェアラブルセンサによる行動認識について，その前提知識を整理した．本章では，本論文がアプローチとして利用するディープネットによる表現学習について説明した後にウェアラブルセンサによる行動認識に対して表現学習を適用する場合の論点について改めて整理し，4 章以降の実験の位置づけについて説明する．

本章では，まずはじめに表現学習およびに深層学習に関する一般的な知識を整理する．次に，深層学習を本研究と同様のウェアラブルセンサによる行動認識問題に対して利用した研究について説明する．これらを踏まえて，ウェアラブルセンサによる行動認識のために有用な表現学習技術を開発するための本研究の論点について説明し，4 章以降の実験の位置づけについて説明する．最後に，以降の章で利用するディープネットおよびに周辺の技術要素に関する前提知識を整理し，各研究への導入とする．

3.1　表現学習と深層学習

Bengio らが指摘するように，機械学習の各種手法の成否はデータをどう表現するかに依存している [10]．2.2 節で述べたように，従来このデータを表現するかという問題は，研究者やアプリケーション開発者が探索的に設計することで行われていた．

表現学習 (representation learning) は，データをどう表現するかを人間の探索によって決めるのではなくデータから獲得するというアプローチである．表現学習は極めて概念的であり，その定義は難しいものの，本論文では「分類器や予測器を構築しやすくするようにデータから表現を学習すること」を表現学習として定義する．これは Bengio らの表現学習における定義と同じである [10]．

類似した言葉として，深層学習 (deep learning) があるが，深層学習とは深さ（あるいは階層性）という多くの領域で重要となる表現に関するヒューリスティックを

その学習に利用した表現学習として捉えられる．すなわち，深層学習とはデータから表現を学習する上での効率的な手法の 1 つとして位置づけられる．

定義からわかるように，表現学習の目的はその後に行う分類器や予測器の学習をしやすくすることであるから，認識精度を高める表現がいい表現であると言える．しかし，これはその認識精度をいかに高めるかということに対する情報は与えていない．適切な表現を学習するという観点にたてば，認識精度を高めるような表現がどのようなものであるかが重要である．Bengio らは，これが網羅するものではないとした上で，一般に良い表現を学習する際に役に立つ prior として次のようなものがあると指摘している [10].

1. 滑らかさ (smoothness)
2. 線形性 (linearity)
3. 複数の説明変数 (multiple explanatory factors
4. 因子 (causal factors)
5. 深さ，あるいは説明変数の階層構造 (depth, or a hierarchical organization of explanatory factors)
6. タスク間での因子の共有 (shared factors across tasks)
7. 多様体 (manifolds)
8. 自然なクラスタリング (natural clustering)
9. 時空間的なコヒーレンス (temporal and spatial coherence)
10. スパースさ (sparcity)
11. 因子の依存関係の単純さ (simplicity of factor dependencies)

この中で，深層学習に最も関係が深いのが深さであり，さまざまな領域においてその有用性が検証されている．深い構造の有用性が示されている領域の例としては次のようなものがある．

画像認識関連

画像認識は初期の頃から深層学習の主要な応用先であり，現在も引き続き最も研究が活発な領域である．深層学習が注目されるきっかけになった Hinton らの研究 [68] は，MNIST における認識精度において当時最高精度であった SVM を破ったものであった．

より複雑な例として，自然画像の認識においても 2012 年に ImageNet データセットを使ったコンペティションである ISLVRC でこれまでの 26.1%のエラー率を 15.3%まで低下させた [13] のを皮切りに，2015 年には同タスクで人間を上回るエラー率を記録した [15]．その他にも顔画像の識別においては 2014 年に 97.35%の精度（それまでの認識精度の最高精度を 27%以上改良）するなど，人間とほぼ同程度の認識精度を達成している [14]．あるいは，本稿でも利用する敵対的訓練を利用した研究事例として，深層学習と敵対的訓練を組み合わせることにより画像を生成する研究が行われている [69, 70, 71]．画像の生成という，これまで難しかったタスクに関しても深層学習の利用によりその可能性が開けている．

音声認識関連

音声認識において，従来的な 3 層のニューラルネット (multi layer perceptron) は音声認識に古くから利用されてきたが，通常その認識精度は隠れマルコフモデル (Hidden Markov Model) とガウシアン混合モデル (Gaussian Mixture Model, GMM) を組み合わせた HMM-GMM と比較すると見劣りするものであった．近年，より深い構造を持つディープネットを GMM の代わりに利用することで単音認識や大規模コーパスを利用した音声認識で GMM を上回る認識精度が達成可能であることがさまざまな研究により報告されている [72, 16, 73, 74, 75, 76]

音声を単純に認識する以外にも，音声合成の分野で深層学習が利用されている [77, 78, 79, 80]．この分野で最も優れた研究結果として結果として，Google の Deep Mind が開発した WaveNet は，音声信号をモデル化することによりアメリカ英語と中国語の音声を合成するタスクで既存の手法を大幅に改良した．具体的には，合成した音声が自然かどうかに関する主観的な評価で人間の発話との差を既存手法比較で 50%以上も改良し，自然な発話や楽曲などを合成することができたことを報告している．合成されたサンプル事例はウェブサイト [1] にて再生することができる．

[1] https://deepmind.com/blog/wavenet-generative-model-raw-audio/

また，マイクロソフトは 2012 年に発表した MAVIS (Microsoft Audio Video Indexing Service) に深層学習を利用するなど，実社会での利用もすでに行われている．

ゲームエージェント関連

従来良いゲームエージェントの設計は，ゲームに有効な特徴を人間が機械可読な形で設計することが必要でありさまざまなゲームごとに異なるエージェントを作る必要があった．Mnih らが発表した論文 [17] では，画像認識と強化学習 (より具体的には Q 学習) を組み合わせることにより人間による特徴量の設計を介さずに Atari の複数のゲームにおいて最もうまい人間を上回るゲームスコアを達成できたを報告している．

ゲームエージェントとして代表的なもう 1 つの例は囲碁に対して利用した例，いわゆる Alpha-Go と呼ばれるものがある [18]．従来囲碁は，その探索空間の広さ (約 10^{360} 程度) や局面を評価する関数（評価関数，value function）設計の難しさから人間のプロレベルには及ばない状況であった．Silver らは，深層畳み込みニューラルネットと自己対局，およびに探索を組み合わせることによりヨーロッパのトッププロである Fan Hui に勝利したことを Nature の論文で発表した [18]．また，2016 年 3 月には韓国のプロと対局し，5 番勝負で 4-1 で Alpha-Go が勝利した．

転移学習関連

これまでに説明した領域と明確に切り分けられる領域ではないものの，深層学習を利用した研究が活発な領域の 1 つに転移学習 (transfer Llarning) がある．転移学習とは，別のタスクで獲得した知識（ニューラルネットで言えば重みやバイアスなどのパラメタ）を新しいタスクに利用する枠組みである．転移学習自体の調査論文としては，Pan らのもの [81] が詳しい．特に，ニューラルネットでは事前に学習したネットワークの重みを再学習 (Fine-tuning) したり，あるいは複数のタスクを同時に解くマルチタスク学習の形で実現されることが多い．また，入力分布の違いを修正するドメイン適応の文脈においても深層学習が有効であることが報告されている [82, 83]．

これらの分野はいずれも，特徴量の獲得に課題があった分野である．例えば，画像認識や音声認識においては高次元な生データをどのように表現するかが課題であっ

た．また，ゲームエージェントにおいても，例えば Atari のようなゲームではエージェントがその画面の状態を自分自身では理解することが（すなわち適切に表現することが）できず，囲碁では盤面の良さをロバストに評価することができないことが問題であった．転移学習も，他のタスクでも利用できるほど汎用性の高い表現を獲得できないことが問題の 1 つであった．本論文では，同様に表現の獲得に課題を抱えるウェアラブルセンサを利用した行動認識タスクに深層学習を適用しその有用性を検証するとともに，ウェアラブルセンサを利用した行動認識アプリケーション上の問題を解消するための深層学習をベースにした手法を提案する．

3.2 ウェアラブルセンサを利用した行動認識に表現学習を適用した既存研究

ウェアラブルセンサを利用した行動認識に対する表現学習の利用は，2011 年に Plötz らが初めて提案した [12] のを皮切りに，特にニューラルネットの構造についての比較が進められてきた．本節では，表現学習を同種のタスクに適用した事例について説明し，本論文の位置づけについて説明する．

Plötz らは，ユビキタスセンサ（ウェアラブルセンサや環境埋込み型センサ）により計測された値から行動を推定する問題において，伝統的な方法の問題点が特徴量をドメイン知識を利用して設計していることであると指摘し，初めて表現学習を利用することを提案した [12]．Plötz らは，1) 複数のタスクに対して汎用的であること，2) ラベルデータに依存しないこと，3) 大規模データセットによる利点を得られること (一方で大規模性に過度に依存しないこと)，4) 本質的な情報を提供すること，5) 実時間の計測に適用可能であること，の 5 つの設計方針のもと，主成分分析と，中間層を 2 つもつ制限付きボルツマンマシン (Ristricted Boltzman Machine) を利用した手法を提案した．提案した表現学習による方法と従来的な特徴量（時間統計量及びに FFT 係数）の認識精度を 4 つの異なるデータセットで比較し，従来の特徴量と比較して認識精度が改善することを報告している．

Vollmer らは，Plötz らと同様の枠組みの中で教師なし学習により表現を獲得する方法として，学習される表現がスパースになるような制約をかけることを提案している [84]．Plötz らと同じ 4 つのデータセットで提案手法を評価し，3/4 のデータセットでスパース性の制約により認識精度が改善することを報告している．

Plötz らや Vollmer らの研究が教師なし学習をベースにした表現学習手法であった

のに対して，Yang らは教師あり学習の手法である畳み込みニューラルネットを利用した表現学習手法を提案した [19]．Yang らは，畳み込み処理により得られる局所的な特徴やプーリング処理により得られる移動不変性，およびにそれらの積み重ねにより得られるさまざまなスケールの特徴が行動認識において重要であることを強調し，主に画像認識で利用されてきた畳み込みニューラルネットを D 次元の時系列データから成るセンサデータに対して適用する方法を提案しその有効性を検証した．具体的には，1) 入力層，2) 畳み込み層（畳み込み処理，プーリング処理，およびにバッチ正規化）の繰り返しを 3 層，3) センサを統合する全結合層，4) 出力層の合計 6 層からなる畳み込みニューラルネットを利用し，時間特徴量，制限付きボルツマンマシンと比較し，提案手法により獲得される表現が高い認識精度を持つことを報告している．類似した研究として，Zhng らも異なるデータセットにおける畳み込みニューラルネットを利用した際の有効性を報告している [85]．

教師あり学習を基本とした表現学習手法に関する別の研究として，Hammerla らは 4,000 回以上におよぶ大規模な実験を通してニューラルネットの構造およびにハイパーパラメターが認識精度に与える影響を調査した [21]．具体的には，ニューラルネットの構造として単純な全結合層を重ねた順伝播型ニューラルネット，Yang らが利用したのと同様の畳み込みニューラルネット，時系列性を考慮したニューラルネットである LSTM (Long Short Term Memory) およびに b-LSTM (Bi-Directional LSTM) を検討し，各モデルごとに学習率や正則化方法，モデル構造が認識精度に与える影響を評価した．評価には 3 つのデータセットを利用し，1) 認識精度の最大パフォーマンスでは多くの場合時系列性を考慮した方法 (LSTM または b-LSTM) が良い認識精度となること，2) パラメタに対する安定性では畳み込みニューラルネットが優れていること，3) 重要なパラメタはモデルにより異なることなどを報告している．

Hammerla らと同様に再帰的な構造を取り入れた研究として，Ordóñez らは畳み込み処理と再帰的処理を組み合わせた Convolutional LSTM を利用することを提案し，Convolutional LSTM が CNN や LSTM を上回る認識精度を達成したと報告している．[20]．

より応用に近い研究として，Hammerla らは，実環境において得られたパーキンソン病の病状評価に深層学習を利用することを提案している [86]．具体的には制限付きボルツマンマシンを利用したモデルと従来的な特徴量による認識精度を 34 名のパーキンソン病患者のデータセットを利用した実験により比較し，深層学習が実環境において得られる信頼度の低いラベルデータを利用する場合でも既存の手法を上

回る認識精度を達成可能であることを報告している．また，Sathyanarayana らは，加速度データから睡眠の質を予測するタスクに深層学習を利用することを提案している [87]．具体的には，畳み込みニューラルネットと従来的な手法 (ロジスティック回帰，SVM，ランダムフォレスト，アダブースティング) を比較し，畳み込みニューラルネットが state-of-the-art な手法の認識性能を 8%程度改善すること，また畳み込みニューラルネットの場合には従来利用されてきた前処理技術を利用しなくても睡眠の質を正しく予測できることなどを報告している．

以上まとめると，これまでの研究により次のようなことが明らかになっている．

- 教師なし (制限付きボルツマンマシン) およびに教師あり (CNN, LSTM, b-LSTM, Convolutional LSTM) の深層学習による方法が，従来利用されてきた特徴量と比較して良い認識精度の達成に寄与しうること [12, 84, 19, 85, 21, 20, 86, 87]．
- 単純な歩行運動 (歩く/走る/寝る等) から睡眠の質，パーキンソン病の病状といった様々なデータセットにおいて良い特徴を獲得できる可能性があること [19, 86, 87]．
- 教師データを活用可能な場合には，教師データを利用したモデル (特に CNN や LSTM およびにその亜種) が良い精度を達成すること [19, 21, 20]．
- 実世界で得られるユーザからの行動報告という信頼度の低いラベルデータを利用する場合でも (教師なし) 深層学習による特徴抽出は機能すること [86]．

3.3 本研究の論点と 4 章以降の位置づけ

さて，2 章ではウェアラブルセンサを利用した行動認識においてラベルデータがないあるいは少ないユーザに対して認識精度を高めることおよびにおよびにプライバシーへの配慮をいかに行うかがウェアラブルセンサを利用した行動認識を行う上で課題となること，およびに既存の研究ではこの問題に対しての検討及びに提案が不十分であることを確認した．また，前節ではウェアラブルセンサによる行動認識問題において深層学習はすでに利用され始めており，またその有効性はさまざまな行動を対象とする場合において検証されていることを確認した．一方，深層学習の有用性が認識され始めたのはここ 5 年余であり，実用上重要となるユーザ汎用性やプライバシー保護に対する議論に至っていない．

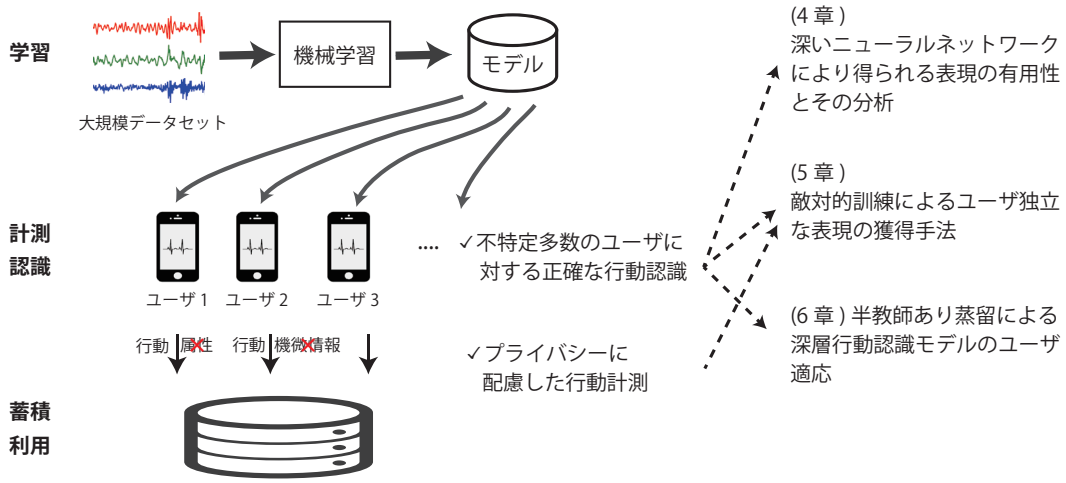


図 3.1: 本論文の論点と以降の章の対応関係.

以上をまとめるに，ウェアラブルセンサを利用した行動認識のための表現学習技術の実現に向けた本論文の重要な論点は次の 2 点である．本論文ではこの 2 つの論点に対する検討及びに技術的な手法の提案を行うことで，ウェアラブルセンサによる行動認識に対して表現学習が有用であることを示す.

本論文の重要な論点

1. 不特定ユーザに対する高精度な行動認識のための表現の獲得問題
2. プライバシーに配慮した表現の獲得問題

以降 4 章から 6 章とこれらの論点の関係を整理し，本論文における各章の位置づけを明確化する．先回りしてまとめると，これら 2 つの論点と以降の章の対応関係は図 3.1 の通りである.

4 章では，深いニューラルネットにより獲得される表現と従来利用してきた特徴量とのユーザ独立認識の問題設定における比較を行う．具体的には，ウェアラブルセンサを利用した 4 つのデータセット合計 110 万サンプル以上の教師付きデータに対して最大 24 層のニューラルネットを適用し，ウェアラブルセンサによる行動認識に対してディープネットによる表現学習が有効であることを示し，ウェアラブルセンサによる表現学習技術を発展させることの妥当性を検証する．また，認識精度およびにプライバシーへの配慮の観点から学習された表現を分析することで，5 章 6 章で提案する手法の土台となる知見を確認する章として位置づけられる.

すでに述べたように，ディープネットのウェアラブルセンサによる行動認識に対する有効性はいくつかの研究で確認されてきている．しかし，2.3 節で述べた問題設

定の類型に即して考えると，いずれの研究もユーザ依存認識の問題設定においての有効性の検証にとどまっており，本研究が扱うような実際環境で現れる学習データが少ないあるいは全く得られないユーザを含むような不特定ユーザに対する認識精度という観点からは不十分である．4 章は，既存の研究が示してきたディープネットの有効性を，これまで検証が不十分な問題設定においても検証することでウェアラブルセンサによるる行動認識における表現学習の有効性をより幅広く検証するものである．

5 章では，ディープネットが学習する表現がユーザ独立になるように制約をかける方法，ユーザ敵対的ニューラルネット (User-Adversarial Neural Networks) を提案する．ユーザ独立な表現とは，形式的には表現 R がユーザ変数 U と独立であることを意味しており，具体的にはユーザの違いに影響されないような特徴量を獲得することと対応する．ユーザ独立性は 4 章で述べるような学習データが得られないユーザに対しても汎用な認識を行うために重要となりうることに加えて，センサにより含まれるユーザあるいはその背後にある属性情報を極力取得しないようにするという意味でプライバシー問題に対しても重要である．提案手法は，学習時に表現 R からユーザ U を分類しようとするような敵対的ユーザ分類器を考慮した学習を行うことで，表現 R が持つ認識したい行動に関する情報を最大化しながらユーザに関する情報を最小化するようにディープネットを拡張するものである．3 つの行動分類タスクにおいて提案手法と通常のディープネットの認識精度およびにユーザ独立性の比較を行い，提案手法によりユーザの独立性の高い表現が獲得できることを定量的・定性的双方の観点から確認する．

6 章では，行動認識精度を高めることを目的に，ディープネットのユーザ適応手法，半教師あり蒸留の提案を行う．ユーザ適応は学習データが得にくいユーザに対してその他のユーザから学習された知識を利用することで認識精度を高める試みであるから，適用に利用可能なデータは一般に少数のラベル付きデータや教師なしデータに限定される．そのため，ディープネットをユーザ適応する場合その大量のパラメタを限られたデータを利用して適切に修正するかが問題となる．提案手法である半教師あり蒸留とは，学習済みモデルから得られる情報とラベル無しデータを有効活用することで効率的なユーザ適応を行う手法である．既存の 8 つのユーザ適応手法と提案手法を認識精度の観点で比較し，提案手法はラベル無しデータを活用することで過学習を抑制しながら適切なユーザ適応が行えることを確認する．

3.4 ニューラルネットに関する前提知識

本章の最後に，以降の章を読み進めるために必要となるディープネット及びに周辺技術に関する前提知識を整理する．

ニューラルネット (あるいは人工ニューラルネット，artificial neural networks) は，高度な情報処理実現を目的に，生物の神経回路網を模倣するような仕組みとして開発されたものである．ニューラルネットの歴史は古く，最も初期の研究は 1940 年代にまで遡る．そのため，ニューラルネットの構造についての検討やその学習方法に関する検討などは数多く行われてきているが，ここでは本稿を理解するのに必要な最小限の知識に絞って説明する．

3.4.1 順伝播型ニューラルネット

順伝播型ニューラルネット (feedforward neural networks) は，図 3.2 に示すように，何らかの入力ベクトル x を受け取りスカラ値を出力するユニット (図中丸記号) と，そのユニットと入力の関係を示す重みベクトル w，バイアスパラメタ b，及びに活性化関数 f によって理解することができる．ここで，入力ベクトル x は例えばある時刻における複数のセンサによる観測値に対応し，その場合図中の k は観測値信号の数 d に対応する．あるいは，入力として 2 章で考えたような多次元観測信号を一定間隔の窓で切り出した多次元サブ時系列を入力として受け取る場合，x はそのサブ時系列を 1 次元上に並べたものになり，その場合 k は窓の長さ $l \times d$ に対応する．

この時，このユニットの出力 z は次式のようになる．

$$z = f(w_1x_1 + w_2x_2 + \cdots + w_kx_k + b) = f(\sum_i w_ix_i + b)$$

ここで，b はバイアス (bias) と呼ばれ，ユニットの出力の基準値を決めるようなパラメタである．また，w は重み (weight) と呼ばれ，入力 x をどのように変換するかを決めるようなパラメタである．詳細は後に述べるが，ニューラルネットではこれらの重み w とバイアス b を与えられた学習データを元に修正することで．学習データに適合する関数を学習する．

また，f は重み w とバイアス b により変換された入力 x を写像する何らかの関数であり，一般に活性化関数 (actiavtion function) と呼ばれる．歴史的には，活性化関数には単調増加する非線形関数が用いられ，ロジスティック関数 $f(u) = 1/(1 + e^{-u})$ やシグモイド関数 $f(u) = \tanh(u)$ と呼ばれるものが古典的には利用されてきた．こ

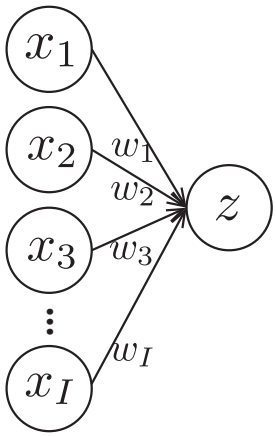


図 3.2: 順伝播型ニューラルネットの基本的な構成要素.

のような非線形関数を利用することにより，重み w やバイアス b による線形変換では捉えられないようなデータの情報を取り出すことができる.

先ほどの図では出力ユニットが z が 1 つの場合で説明したが，より一般的に順伝播型ニューラルネットの各層では複数の値を出力することができる．この時，j 番目の出力を z_j で表すとすると，複数の値が取る場合は次のように一般化して記述することができる.

$$z_j = f(w_{j1}x_1 + w_{j2}x_2 + \cdots + w_{jk}x_k + b_j) = f(\sum_i w_{ji}x_i + b_j)$$

また，より一般的にはベクトル表記を利用して次のように記述される.

$$\mathbf{z} = \mathbf{f}(\mathbf{W}\mathbf{x} + \mathbf{b})$$

また，上記の例では説明のため層の数を 2 層に固定したが，古典的なニューラルネットでは更にもう 1 つの層を追加し，図 3.3 に示すような 3 層構造のネットワークを考える．この時，各層を左から順に $l = 1, 2, 3$ で表し，$\mathbf{W^l}$，$\mathbf{b^l}$，$\mathbf{z^l}$ をそれぞれ重み l 層目の重み，バイアス，出力とすると，$l = 2$ およびに $l = 3$ の出力は次のようになる.

$$\mathbf{z^2} = \mathbf{f}(\mathbf{W^2}\mathbf{x} + \mathbf{b^2})$$

$$\mathbf{z^3} = \mathbf{f}(\mathbf{W^3}\mathbf{z^2} + \mathbf{b^3})$$

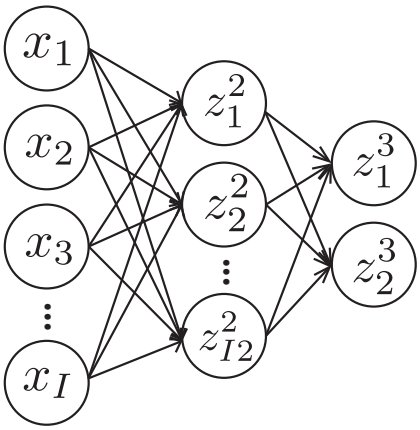


図 3.3: 3 層構造を持つ順伝播型ニューラルネットの模式図.

すなわち，x を 1 番目の層の出力 z^1 と考えると，より一般的に l 層目の出力は

$$\mathbf{z^l} = \mathbf{f}(\mathbf{W^l z^{(l-1)}} + \mathbf{b^l})$$

となり，1 つ前の層の出力を重み $\mathbf{W}$ およびにバイアス $\mathbf{b}$ により線形変換した後に何らかの非線形活性化関数 f により写像することにより得られる．これは，より一般的に L 層構造を持つ順伝播型ニューラルネットで共通である．なお，ここで f は l を引数に持たない形で記しているが，実際には f は共通でなくても良い．

このような処理により得られる順伝播型ニューラルネットの最終的な出力 z^L はネットワークのパラメタ $\mathbf{W}$ およびに $\mathbf{b}$ を変化させると変化する．ニューラルネットでは，何らかの入出力のペアの集合 $\{(\mathbf{x_1}, \mathbf{y_1}), (\mathbf{x_2}, \mathbf{y_2}), \cdots (\mathbf{x_N}, \mathbf{y_N})\}$ が与えられた時に，データ $\mathbf{x_n}$ をネットワークに入力した際の出力 $\mathbf{z^L}(\mathbf{x_n}; \mathbf{W}, \mathbf{b})$ が $\mathbf{x_n}$ と対応する $\mathbf{y_n}$ になるべく近くなるようにパラメタを調整する．すなわち，L を $\mathbf{y_n}$ と $\mathbf{z^L}(\mathbf{x_n}; \mathbf{W}, \mathbf{b})$ の距離を定義する何らかの関数であるとすると，ネットワークは次の最適化問題を解くことにより学習される．

$$\mathbf{W}, \mathbf{b} = \arg\min \sum_{\mathbf{n=1}}^{\mathbf{N}} \mathrm{L}(\mathbf{y_n}, \mathbf{z^L}(\mathbf{x_n}; \mathbf{W}, \mathbf{b}))$$

ネットワークの出力と真の値の距離を図る関数 L は一般に損失関数 (loss function) と呼ばれる．損失関数は対象とする問題によって異なるものの，本論文が主として扱う多クラス分類問題では，交差エントロピー (cross entropy) と呼ばれる損失関数

が利用される．交差エントロピーは次の式により表される．

$$L(\mathbf{W},\mathbf{b}) = -\sum_{\mathbf{n=1}}^{\mathbf{N}}\sum_{\mathbf{k=1}}^{\mathbf{K}} \mathbf{y_{nk}} \log \mathbf{z_k^L}(\mathbf{x_n};\mathbf{W},\mathbf{b})$$

ただし，ここで $y_n = [0001]$ のように対応するクラスが真のクラスである場合にのみ 1 をとりそれ以外は 0 を取るようなベクトルである．また，$z_k^L(x_n;\mathbf{W},\mathbf{b})$ はネットワークの出力層の k 番目のユニットの出力であり，

$$z_k^L(x_n;\mathbf{W},\mathbf{b}) = \frac{\exp(\mathbf{u_k^L})}{\sum_{\mathbf{j=1}}^{\mathbf{K}} \exp(\mathbf{u_j^L})}$$

により求められる値である．ここで分母は出力層のすべてのユニットの出力の総和であり，すなわち正規化項である．この時，$z_k^L(x;\mathbf{W},\mathbf{b})$ は与えられた入力 x が k 番目のクラスに属する確率 $p(y = y_k|x)$ として解釈することができる．すなわち，例えば $z_k^L(x;\mathbf{W},\mathbf{b})$=0.99 の場合には入力 x が k 番目のクラスに属する確率は 99%であり，$z_k^L(x;\mathbf{W},\mathbf{b})$=0.2 の場合には入力 x が k 番目のクラスに属する確率は 20%というように解釈できる．

このような交差エントロピーを最小化することで，ネットワークは訓練データをよく識別するようなモデルを学習することができる．一方，このような損失関数 L は一般に凸関数ではなく，大域的な最小解を解析的に得るのは通常不可能である．そのため，特に多層構造を利用したニューラルネットでは，大域的な最小解を求める代わりに，L に関する局所的な極小点を求めることを考える．

L に関する局所的な極小点を求める多層ニューラルネットで利用される最も代表的な方法に，勾配降下法 (gradient descent) がある．ここで，勾配 (gradient) とは

$$\nabla L \equiv \frac{\partial L}{\partial w} = \left[\frac{\partial L}{\partial w_1} \cdots \frac{\partial L}{\partial w_M}\right]$$

で計算されるベクトルである．勾配降下法では，まず何らかの初期値で w を初期化したのち，現在のパラメタ w を次の式で更新することを繰り返し行う．

$$w^{t+1} = w^t - \eta \nabla L$$

ここで t は繰り返し回数であり，w^t は t ステップ目のパラメタを意味する．また，η は学習率 (learning rate) と呼ばれるパラメタであり，1 回あたりの更新幅を意味する．

3.4.2 ディープネット学習のための工夫

前項で述べたように，多層のニューラルネットを構築するのは比較的単純であり，層を重ねることにより多層モデルを構築するというアイディアは古くから検証されていた．一方，このような深い構造を持つネットワークを学習することは難しく，特に次の 2 つの問題があった．

第 1 の問題は，過学習 (overlearning) あるいは過適合 (overfitting) と呼ばれる現象である．過学習とは，学習するパラメタ θ が訓練に利用するデータに対して過剰に適合してしまい，訓練データに対する誤差 (訓練誤差，training error) と訓練にはまだ現れていない将来予測したいサンプルに対する予測誤差 (汎化誤差，generalization error) とが乖離してしまう現象のことを指す．通常実際に最適化したいものはまだ現れないサンプルに対する汎化性能であるから，過学習は学習されたパラメタの汎用性の観点からして問題である．

過学習の問題は，ネットワークの自由度 (すなわちパラメタの数) が大きいほど生じやすくなる現象である．すなわち，パラメタが多いとその分柔軟にデータに適合できてしまうため，放っておくと必要以上に訓練データに過適合してしまう．一方ディープネットの利点はパラメタの多さと階層構造によるその表現力の高さであり，過学習の問題はディープネットを利用する際の本質的な問題の 1 つと言える．

第 2 の問題は，勾配消失問題 (vanishing gradient problem) である．勾配消失問題とは，出力層と与えられた訓練データにより計算される損失の勾配 ∇L が，多層ネットの特に入力に近い部分ではほとんど 0 になってしまうという問題である．勾配が 0 になるということは，パラメタがほとんど更新されない，すなわちパラメタが初期値からほとんど動かないことを意味する．この問題は，3 層程度の浅いネットワークでは問題にならないが，多層ネットワークでは深刻な問題であり学習を行う上で最大の障壁であった．

これらの問題から学習が難しいとされていたディープネットであるが，近年はいくつかの理由により学習が可能になっている．まず第 1 の理由は，大規模データが整備されたことである．大量の学習データを事前に用意して学習を行うことで，仮にデータに過剰に適合したとしてもそのパラメタが未知のデータに対しても有用である可能性は高まる．例えば，画像認識でよく利用される手書き文字データセットである MNIST では 60,000 件の 0–9 までの手書き文字データが提供されている．また，一般画像認識のためのデータセットである ImageNet では 1000 のカテゴリに対して合計 120 万枚の画像データが提供されている．このような大規模データセット

の整備により，過適合の問題が相対的に小さくなったことが第 1 の理由である．

第 2 の理由は，計算機の計算能力の向上である．GPU (Graphic processing unit) や CPU の進展により，大量のデータを利用した学習を行うことが現実的な環境で可能になった．それにより，今までより大量のデータを利用した実験や，さまざまなパラメタによる実験などこれまでは困難であった実験が可能になり，大量のパラメタを持つディープネットの学習が現実的に行えるようになった．

第 3 の理由が，効率的な学習アルゴリズムやテクニックの開発である．学習アルゴリズムの代表的な例として，現在ではほとんど利用されないものの，深層学習ブームの火付け役となった Hinton らの研究で利用された事前学習 (pre-training) がある [68]．ここでは特に本論文で利用するテクニックについて説明する．

ReLU

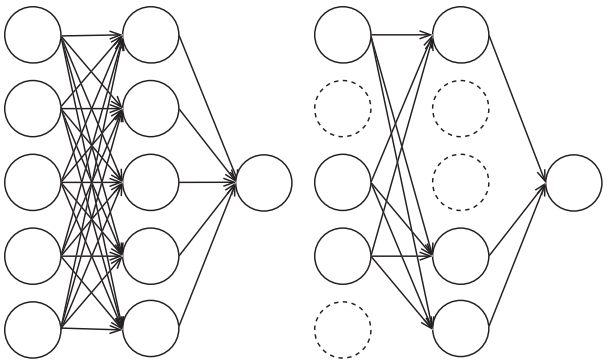
ReLU (Rectified Linear Unit，正規化線形関数) とは，2010 年に Nair らが提案した活性化関数の 1 つである [88]．ReLU は次の式で表される区分線形関数である．

$$f(x) = \max(0, x)$$

LeCun，Bengio、Hinton らが書いた調査論文では，少なくとも 2015 年時点で ReLU による認識精度が一般に高いとされており [89]，学習速度が早く最終的な結果も良いことが知られている．例えば Glorlot らは ReLU を利用することで事前学習なしに良い精度を達成できることを画像およびに自然言語データセットで検証している [90]．

従来のシグモイド関数や双曲線正接関数などを活性化関数 f として利用する場合，f による写像を繰り返し行う，すなわち多層に積み重ねると前述した勾配爆発問題や勾配消失問題が起こる．一方，ReLU は何度重ねても勾配は 0 または 1 であり，原理的に勾配消失問題は起こらない．

行動認識の分野においても，活性化関数としては ReLU が標準的に利用されている [19, 21]．そのため，本稿でも特に言及がない限り活性化関数には ReLU を利用することとする．



(a) ドロップアウト前の状態(b) ドロップアウト中の状態

図 3.4: ドロップアウトの概念図．図中点線で描かれた丸は擬似的に削除されたユニットを指す．

ドロップアウト

ドロップアウト (dropout) とは，2013 年に Hinton らにより提案 [91, 92] されたニューラルネット効率的に正則化する方法の 1 つである．後述する 2016 年に提案されたバッチ正規化の出現により従来より必要性は低下しているものの，発表されてから現在までさまざまな場面で利用されている．ドロップアウトでは，多層ネットワークのユニット (あるいは結合 [93]) を確率的に取り除いて学習を行う方法である．このように学習を行うことにより，特定のユニットが前後の特定のユニットと密に影響しあうことを抑制し，また特定のユニットが存在しない場合でもうまく予測できるような頑健な特徴を学習することができる．

具体的にはまず，学習時には対象とするユニットを事前に設定した確率 p でランダムに擬似的に取り除く (図 3.4)．そして，この全ユニットのうち一部のユニットのみを利用したネットワークを通常の勾配法により学習させる．すなわち，訓練サンプルをネットワークに入力し損失 L を計算し，誤差の勾配を利用してネットワークの重みを更新する．このような手順を各勾配の計算ごとに繰り返し行い学習を行い，ネットワークのパラメタを更新する．

このように学習することにより，過学習を抑制する効果があることが知られている [92]．

早期終了

ReLU やドロップアウトとは違いディープネットの学習に限った方法ではないものの，過学習を起こさないための方法としてよく利用される方法に早期終了 (early stopping) と呼ばれる方法がある．前述したとおり，ニューラルネットに限らず教師あり学習の目標は汎化誤差を最小化することであり，過学習とは学習を進めると訓練誤差は低下するがテスト誤差が増加するような現象のことを指すものである．早期終了とは，学習データとは別に用意した検証データを利用してテスト誤差を計測し続けることで，訓練誤差は低下するがテスト誤差は低下しないような状態に陥った場合に学習を停止させる方法である．

転移学習

ディープネットの学習が近年うまく行くようになった理由の 1 つに，学習データの整備があることはすでに述べた．一方で，実際にニューラルネットを利用したアプリケーションを作成しようとする場合にはそのような大量の学習データを用意することは難しい場合も多い．学習データが限られている場合の対策として，特に画像認識など標準的なデータセットが存在し，また学習済みモデルの情報が公開されている種類のデータを対象とする場合には学習済みモデルを利用した Fine-Tuning がよく利用される．ここで，Fine-tuning とは目的とする問題 P に対応する小規模なデータ D があるときに，大規模な訓練集合 D' で訓練されたパラメタ θ' を問題 P を解くためのニューラルネットの初期値として利用した上で学習を行う，改めて D で学習を行う方法である．

このような Fine-tuning がうまく機能するかどうかは，訓練集合 D' により学習されたパラメタ θ' がどの程度 D に対しても有用であるかによる．すなわち，Fine-tuning は対象としている問題 P と類似したタスクに関する大規模なデータセットが存在する場合に有効である．一方，例えば一般物体認識のデータセットである ImageNet で学習したネットワークを物体検出やセマンティックセグメンテーションなどの領域に利用することで良好な認識精度が得られることを報告しており，学習データが少ない場合の標準的な方法として利用されている．

3.4.3 畳み込みニューラルネット

畳み込みニューラルネット (Convolutional neural networks，CNN) は，画像や系列データなどの構造データに対してうまく働くことが知られている順伝播型ニューラルネットの 1 つである．本来画像認識において提案されたものであるが，畳み込みネットワークが与える移動不変性の性質は多くの時系列構造を持つデータと相性がよく，また，そのため例えば加速度センサを利用した行動認識においてもよく利用されている．ここでは，まず通常の畳み込みニューラルネットについて特に順伝播型ニューラルネットとの違いを説明した後に，ウェアラブルセンサを利用した行動認識タスクで取り扱う多次元系列データへの適用方法を説明する．

畳み込み処理

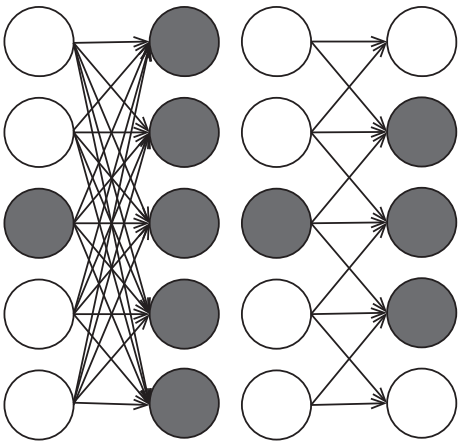
順伝播型ニューラルネットと畳み込みニューラルネットの最も大きな違いは，層の間の結合方法である．図 3.5 は，3.4.1 項で説明した順伝播ネットワークにおける層間の結合方法と畳み込みネットワークの結合方法を比較したものである．このように 3.4.1 項で説明した順伝播ネットワークは隣接した層の間はそれぞれのユニットすべてが全結合 (fully-connected) されたものであるのに対して，畳み込みネットでは特定のユニットのみが結合を持つ形となる．特に，畳み込みネットワークでは出力側のユニットは入力側の特定の範囲のユニット (入力が 1 次元の場合は x_i から x_{i+k}) のみ入力を受けような制約を持つ．なお，以降すべての層が全結合したような順伝播型ニューラルネットのことを畳み込みネットワークと区別して全結合ネットワーク (fully-connected networks) と呼ぶこととする．

畳み込みネットワークの l 層目の j 番目のユニットの出力 z_j^l は次の式により計算される．

$$z_j^l = f((\mathbf{z^{l-1}} * \mathbf{h})(\mathbf{j})) = \sum_{\mathbf{p=0}}^{\mathbf{H-1}} \mathbf{z_{j+p}^{l-1}} \mathbf{h_p} + \mathbf{b_j}$$

この時，$*$ は 2 つのベクトル $\mathbf{z}$ と $\mathbf{h}$ の間に定義される畳み込み処理であり，また $\mathbf{h}$ は入力を変換する重みパラメタに相当するものである．一般に，$\mathbf{h}$ はフィルタ (filter) あるいはカーネル (kernel) と呼ばれる．本書では統一してフィルタと表記することとする．

このような畳み込み処理は，入力中にある特定のパターンを検出する働きを持つ．すなわち，出力 $\mathbf{z_j^l}$ は入力中にフィルタのと類似したパターンが存在する箇所で大き



(a) 全結合層　　(b) 畳み込み層

図 3.5: 全結合層と畳み込み層の層間の結合方法の比較．図中の黒く塗りつぶされた箇所はある入力ユニットがどの範囲に影響をあたえるかを示したものである．

くなる．そのため，このような畳み込み処理により入力中の局所的な構造パターンを検出した出力を与えることができる．

上記の例では最も単純に，1 つのフィルタによる畳み込み処理を考えた．しかし，1 つのフィルタでは 1 つのパターンしか検出することができない．畳み込みニューラルネットでは通常，複数のフィルタを用意して複数の出力を獲得する．また，畳み込みニューラルネットでは複数の特徴マップを入力に受け取ることができる．この時，各フィルタに対応する出力は一般に特徴マップ (feature map) と呼ばれる．また，入力される各特徴マップは入力チャネルと呼ばれ，またその数はチャネル数と表記される．

入力チャネルが K 個，フィルタ数が M 個の畳み込み層の出力は，次の式のようになる．

$$z_{jm}^{l} = f((\mathbf{z^{l-1}} * \mathbf{h_m})(\mathbf{j})) = \sum_{\mathbf{k=0}}^{\mathbf{K-1}} \sum_{\mathbf{p=0}}^{\mathbf{H-1}} \mathbf{z_{j+p,k}^{l-1}} \mathbf{h_{pkm}} + \mathbf{b_{jm}}$$

ただし，$z_{j,k}^{l-1}$ は $l-1$ 層目の k 番目の特徴マップの j 番目のユニットの値を，$\mathbf{h_{km}}$ は k 番目の特徴マップに対する m 番目のフィルタを，b_{jm} は m 番目のフィルタの j 番目のユニットのバイアスをそれぞれ意味する．この時入力チャネル数 K によらず出力のチャネル数は M になる．

畳み込みネットワークにおいても，前述した全結合ネットワークと同様，勾配降下法によるパラメタの最適化が行われる．この時最適化されるパラメタは M 個のフィルタの重み $\mathbf{h}$ とバイアス $\mathbf{b}$ となる．

このような畳み込み処理が行動認識において有用と思われる理由は大別して次の2つである．第1に，スライディングウィンドウに伴う開始位置の任意性である．行動認識で利用されるスライディングウィンドウにより与えられる入力 $\mathbf{x}$ は，どの位置で切り出しを行うかによりその時間方向のどの位置の情報を含むかに任意性がある．そのため，全結合ネットワークのように各次元それぞれに対して別の重みによる変換を行うのではなく同じフィルタによる変換を行うことが必要となる．第2に，行動の構成性およびに局所性である．例えば，歩くという行動を例に取ると，歩くという行動は足をあげる，下げると言ったより下位の短時間に起こる行動の組み合わせにより表される．すなわち，歩くという行動を認識することはその下位行動の検出とその並びから上位の行動を認識することと捉えることができる．このように，行動はより短時間の行動素 (action primitive) により構成され，各行動素と対応する信号は観測信号中のさまざまな場所に局所的に存在していると考えられる．そのため，局所的なパターンを抽出する畳み込み処理との相性が良いと考えられる．

プーリング

畳み込みネットワークで利用されるもう1つの特徴的な処理に，プーリング (pooling) 処理がある．プーリング処理は視覚野における複雑型細胞 (complex cell) の挙動のモデルとなっており，畳み込みネットワークの特徴である移動不変性を与える重要な役割を持っている．

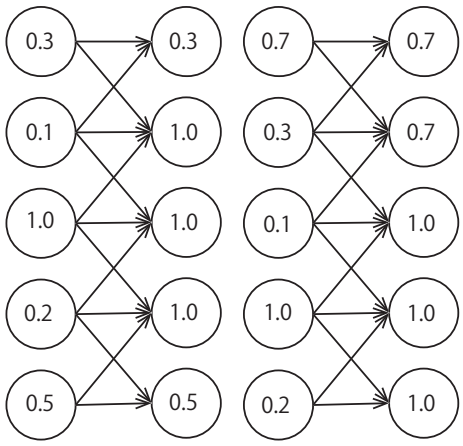
プーリング処理では，入力情報の解像度を低下させることにより入力のずれに対する不変性を与えるものである．例えば，最大プーリング (max pooling) では入力 z^l_{jm} について次のように一定範囲内の最大値を出力することで解像度を低下させる

$$v^l_{jm} = \max_{0 \le q \le Q-1} v^l_{(j+p)m}$$

ここで，Q はプーリングの幅であり，最大プーリングでは範囲内にある Q 個出力の最大値のみを情報として残しその他の情報を欠落させる．

その他によく利用される方法には一定領域内の平均値を出力する平均プーリング (average pooling) や，それらを一般化した Lp プーリング (Lp pooling) がある．本論文ではプーリングについては変更せず最大プーリングを利用することとする．また，以降単にプーリングとよんだ場合には最大プーリングを指すこととする．

プーリングを行うことは移動不変性と密接な関係がある．例えば，入力の系列データ $\mathbf{x} = [\mathbf{x_1 x_2} \cdots \mathbf{x_I}]$ と，$\mathbf{x}$ を1ビット右にずらしたような観測 $\mathbf{x}' = [\mathbf{x_0}, \mathbf{x_1 x_2} \cdots \mathbf{x_{I-1}}]$



(a) $\mathbf{x}$ を入力にした場合　(b) $\mathbf{x}'$ を入力にした場合

図 3.6: プーリングによる移動不変性の概念図．各図中左側が畳み込み処理による出力ユニットを，右側が最大プーリングの結果を示したものである．

の 2 つを考える．これは例えば，画像を例に取れば 1 ビット上下左右いずれかにずらした場合に対応し，実質的には同一の意味持つものとして扱われることが妥当と思われる．また，前述した行動認識で標準的に利用されるスライディングウィンドウ方式では，切り方は任意でありその意味で取り出される系列の開始位置には任意性がある．そのため，このような 2 つのベクトルを実質的に同じものとして扱えることが望ましい．

一方，このような 2 つのベクトルに対して，単なる畳み込み処理では 2 つの入力を同様のものとして扱うことはできない．すなわち，何らかのフィルタ $\mathbf{h}$ を考えた時に，$\mathbf{x}$ と $\mathbf{x}'$ を入力にした場合の出力 $\mathbf{z}$ と $\mathbf{z}'$ はやはりその両端を無視すればやはり 1 ビットずれたようなものになる．また，全結合ネットワークでは，各入力次元をそれぞれ独立のものとして扱うため，出力は全く異なるものになる可能性がある．

一方プーリングでは解像度を低下させることで，$\mathbf{x}$ と $\mathbf{x}'$ をより類似したものとして扱うことができる．そのことを模式的に示したものが図 3.6 である．この例では各プーリングユニットはそれぞれ 3 つの入力ユニットと接続されている．すなわちプーリング幅 $Q = 3$ である．この例に示すように，入力される値がすべて違う (この場合は横に 1 つずれている) 場合でも，プーリングにより出力される値のうち 2/5 は同じになる．

このような移動不変性はウェアラブルセンサを利用した行動認識においても重要である．この理由は，実世界のデータを観測した信号の系列 $\mathbf{r}$ を考えた時に，ある

行動 a がどのタイミングで発生するのかには任意性がある．そのため，行動認識において利用されるスライディングウィンドウが与える入力において，観測された信号 $\mathbf{x}$ のどの位置から行動 a が開始するかはランダム性がある．そのため，行動認識では時間方向へのずれに対する不変性が重要となる．

畳み込み処理の行動認識における利用方法

さて，ここまで畳み込みネットワークを構成する基本的な要素である畳み込み処理とプーリング処理に説明してきた．最後に，この畳み込みネットワークをウェアラブルセンサによる行動認識における入力に適用する方法について説明する．以降，入力データ $\mathbf{X}$ として，スライディングウィンドウにより切り出された D 次元の多次元系列 $\mathbf{S_k}$ を考える．ここで，各次元データはそれぞれ異なるセンサの観測値の系列に対応し，すなわち D 個の観測信号を持つ．例えばセンサとして 3 軸加速度センサを利用する場合には $D = 3$ になり，加えて 3 軸ジャイロスコープを利用する場合には $D = 6$ である．また，各系列はそれぞれ長さ L であるとする．この時，各観測信号をどのように扱うかにより次の 3 つの畳み込み処理が考えられる．

第 1 に，画像認識が RGB 情報を入力チャネルとみなすのと同様に，各系列データをそれぞれ入力チャネルとみなす方法である．一方，画像の RGB 情報が画像という実質的に 1 つの情報を 3 つのチャネルで表しているのに対して，ウェアラブルセンサによる入力は加速度やジャイロなど異なる情報を表しているものである．そのため，筆者が知るかぎりこの方法は利用されていない．

第 2 に，各系列ごとに別々にフィルタ h_d を用意して各系列ごとに畳み込み処理を行う方法である．定式的に書くと，この場合の畳み込み処理の出力は次のように計算される．

$$\mathbf{z_m^d} = \mathbf{f}(\sum_{\mathbf{k=0}}^{\mathbf{K-1}}(\mathbf{x_k^d} * \mathbf{h_{km}^d}))$$

この方法は，Zang らにより利用されている [85]．

第 3 に，すべての系列で共有のフィルタ h を用意して各系列ごとに畳み込み処理を行う方法である．定式的に書くと，この場合の畳み込み処理の出力は次のように計算される．

$$\mathbf{z_m^d} = \mathbf{f}(\sum_{\mathbf{k=0}}^{\mathbf{K-1}}(\mathbf{x_k^d} * \mathbf{h_{km}}))$$

第 2 の方法との違いは，第 2 の方法が観測信号ごとにフィルタ $\mathbf{h_{km}^d}$ を利用しているのに対して，第 3 の方法ではすべてに共通のフィルタを $\mathbf{h_{km}}$ 利用している点である．

この方法は，Jian らや Hammerla らにより利用されている [19, 21].

筆者が知る限り，このいずれの畳み込み処理が適しているかに言及したり，あるいは検証している文献は存在しない．また，いずれの方法が適しているかも自明ではない．ただし，本稿ではこの問題は扱わずにすべての実験で第 3 の方法を利用する．この理由は次の 2 点である．第 1 に，第 1 の方法は第 1 層目で D 個の観測系列が混在した特徴マップへと変換されてしまうため，各系列の特徴が十分に抽出できない可能性がある．第 2 の理由は，第 2 の方法は第 3 の方法と比較すると D 倍のパラメタが必要であるため全体のパラメタ量が増加することである．また，それぞれのセンサは系列データであり，重要なパターンには何らかの共通性があり，各々別のフィルタを用意することは効率が良くないと考える．そのため，以降畳み込みネットワークをウェアラブルセンサによるデータに適用する場合，暗黙的に第 3 の方法で畳み込み処理を行ったことを指すものとする．

3.4.4 蒸留

本論文で提案する手法の一つである半教師あり蒸留によるユーザ適応手法 (6 章) は，Hinton らが提案した蒸留 (distillation) [94] をユーザ適応の文脈に拡張したものである．Hinton らは蒸留をアンサンブルモデルの小型化に利用することを提案し，ドロップアウトを利用したアンサンブルニューラルネットを小型化しながら認識精度を維持することができることを実験的に検証した．また，アンサンブルモデルに対する適用のみにとどまらず，現在では蒸留を元にした手法としてより望ましい構造を持つニューラルネットに変換する手法や [95]，複数のニューラルネットが学習した知識を組み合わせる [96] ためにも利用されている

蒸留とは，何らかの学習済みモデルが存在するときに，学習済みモデルの出力を模倣するように新たなネットワークを学習させる方法である．すなわち，単純なネットワークの学習と比べると，今までになかった模倣プロセスが追加されている．この時，蒸留の文脈では前者を教師モデル，後者を生徒モデルと呼ぶ．これは，前者のネットワークがすでに学んだことを後者のネットワークに教える形を取ること示すものである．

形式的には，蒸留は次の 2 つのステップにより表される．

ステップ **1:** 教師モデルの学習

教師データがついたデータセット D を利用し，大量のパラメタを持つモデル f_t を学習する．定式的には，

$$f_t = \arg\min_f \frac{1}{N}\sum_{i=1}^{N} \mathrm{L}(y_i, \sigma(f(X_i))) + \Omega(||f_t||) \tag{3.1}$$

を最適化する．第 2 項目はなんらかの正則化項であり，例えばドロップアウトや L2 正則化などが該当する．

ステップ **2:** 生徒モデルの学習

学習した f_t とデータセット D' を利用して，モデル f_s を再度学習する．定式的には，

$$\begin{aligned} f_s = \arg\min_f \frac{1}{N'}\sum_{i=1}^{N'} [\lambda\, \mathrm{L}(y_i, \sigma(f_s(X_i))) \\ + (1-\lambda)\, \mathrm{L}(s_i, \sigma(f_s(X_i)))] \end{aligned} \tag{3.2}$$

を最適化する．ここで，s_j^i はソフトターゲットと呼ばれ，

$$s_i = \sigma(\frac{f_t(x_i)}{T}) \tag{3.3}$$

で計算される，教師モデルの出力層の活性状態を温度 T により滑らかにしたものである．また，λ はどの程度教師を強く模倣するかを制御するパラメタである．

式 3.2 にあるように，蒸留は生徒モデルが通常の入力出力のペアから定義される損失の度合い (第 1 項) に加えて教師モデルの出力を模倣するような正則化 (第 2 項) を加えたものとして解釈することができる．例えば，$T = 1$ のとき s_i は i 番目のサンプルを入力にした際の教師モデル f_t の出力 (ソフトマックス関数がかかったもの) であり，第 2 項目は f_t と f_s の出力確率がなるべく近くなるように正則化しているものとみなせる．このように学習することは一見無意味なようにも思えるが，例えば f_s の持つパラメタが f_t よりも小さい場合には出力を模倣する項を入れることにより効率的に学習できることが知られている．また，教師モデルに深い構造を持つモデルを利用して生徒モデルに浅い構造を持つモデルを利用することで，深い構造と同程度の精度を達成できることが報告されている [97].

第4章　ディープネットによるユーザ独立認識と学習された表現の分析

4.1　ユーザ独立認識とディープネット

本章では深い構造を持つニューラルネットをユーザ独立認識問題に対して適用し，その有効性およびに問題点を検証する．これまで述べたように，ディープネットは，画像認識や音声認識で特に高い認識精度を発揮する基礎になっている技術であり，人間による特徴設計を介さずに生のデータから何らかの損失関数を最適化するような特徴を直接取り出す点に特徴がある．ディープネットがユーザ独立行動認識において重要になりうると考える理由は大別して次の 2 点である．第 1 に，ディープネットは損失関数を最適化することにより特徴量を抽出するため，人による判断を含まない点である．慣性センサのように生に近い高次元のデータの特徴を人間が視覚的に発見したり，あるいは機械が扱える形に数値化したりすることは困難である．そのため，人により設計する特徴量で利用できる特徴量が比較的簡易なものに限られるのに対して，ディープネットを利用する場合にはユーザに依存しないような特徴量が複雑であった場合でも抽出できる可能性がある．また，不特定多数のユーザのデータを学習に利用した場合，特にモデルの表現力を限定すれば学習データを効率的に認識するためには不特定多数のユーザに共通の特徴，すなわちユーザに影響されない特徴量を利用することが必要となる．

第 2 に，慣性センサとの親和性である．一方，ユーザ依存認識において，ディープネットは慣性センサデータから行動を識別するような特徴をうまく抽出できることが報告されている [12, 19, 21]．後に見るようにユーザ依存認識においてよく識別する特徴量がユーザ独立認識においても同様に良く認識するかは自明ではないものの，扱うセンサという観点で両者は同一であり少なくともユーザ独立行動認識に対しても良い精度を達成する蓋然性が高い．これらの理由から，ディープネットが既存の特徴設計と比較してユーザ独立行動認識を行う上で優れた特徴量を獲得することが期待される．

まずはじめに，4つの行動認識データセット（合計35名，110万サンプル）のデータに対して深いニューラルネット（6層から最大24層）を適用し，経験的に利用されてきた特徴量と学習された表現の認識精度を比較する．また，ユーザ依存性の観点から学習された表現の分析を行い，単純な深いニューラルネットの適用の限界およびにその問題点について議論する．本章はユーザ独立行動認識に対するディープネットの有効性に関する基礎分析を通して，手法改良への足がかりとなるものである．

本論文に対する本章の貢献およびに得られた知見をまとめると次のようになる．

- 大規模なデータセットを利用して，ディープネットの一種であるCNNにより学習される特徴量と行動認識の文脈で経験的に利用されてきた特徴量とをユーザ独立認識精度の観点で定量的に比較し，CNNによりユーザ独立認識精度を有意に高められることを確認した．この結果は，認識精度を高めたいアプリケーション提供者にとって参考になるものである．

- 一方で，新規ユーザに対する認識精度は学習時に含まれるユーザに対する認識精度と比較すると顕著に低下していた（F値で最大0.30ポイント低下）．また，学習された特徴量の可視化結果は，特徴量がユーザに関する情報を保持したものであることを示唆していた．これらの結果は，ユーザ独立認識を正確に行うために単純なディープネットの適用を超えた手法の開発を動機づけるものである．

- より層数の多いニューラルネットを利用した場合，ユーザ依存認識精度は有意に高まる一方で，ユーザ独立認識精度は有意に変化しないか，あるいは低下する場合もあった．この結果は，どちらの認識を行うのかにより最適なモデル構造は異なること，また行動認識研究ではどちらの認識を対象としているのかを明示する必要があることを示唆している．

- 認識精度の低下度合いは行動クラスにより大きく異なっていた．例えば，JumpingやRunnning Fowrardといった特徴のはっきりした行動は認識精度の低下が少なく（0.10ポイント以下），Walking Forward（WF），Walking Upstairs（WU）など類似した行動間で間違いが増加する傾向にあった．この結果は，複雑なクラス分類を行おうとするほどユーザ独立認識が難しくなることを示唆している．

4.2 利用データセット・ネットワーク構造

4.2.1 データセット

ユーザ独立行動認識における深いネットワーク構造の有効性およびに問題点の検証にあたり，ウェアラブルセンサを利用した行動認識データセットを利用した実験を行った．データセットとしては，可能な限り一般的に利用可能な知見を得るために，利用目的やデータ量が異なる次の 4 つを利用した．

1 つ目のデータセットとして，Opportunity Recognition Dataset[1]（以下 Opp と表記する）[98, 99] を利用した．Opp 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．

Heterogeneous Activity Recognition Dataset (以降 Hetero と表記する）は，計測に利用するデバイスの種類の違いが認識精度に及ぼす影響を調査する目的で 2015 年に公開された比較的新しいデータセットである [100]．決められた行動スクリプトを実行する 9 人のユーザの行動を加速度とジャイロスコープにより計測したデータとなっている．行動としては，Biking, Sitting, Standing, Walking, Stair Up and Stair down の 6 つの行動に加えてそれ以外の合計 7 種類が提供されている．

USC-HAD (以降 USC と表記する) は，公開されている行動認識データセットの中では比較的多い男女 7 名ずつ合計 14 名のデータを含む点が特徴である．行動としては，前に歩く，右に曲がる，左に曲がるなどの一般的な生活における最も基本的な 12 種類の行動が記録されている[3]．データの計測は MotionNode [4] により行われており，6 次元の系列データである．

Wheelchair データセット (以降 Wheel と表記する) は，9 名の車いすユーザの行動を記録したデータセットである．9 名の内訳は，男性 7 名女性 2 名，6 名の手動車いすユーザおよびに 3 名の電動車いすユーザ，20 歳から 60 歳と幅広い属性を持つデータセットである．本データセットは，車いす行動により都市のアクセシビリティを推定する試みに向けて作成されたデータセットであり，1) 段差を乗り越える/乗り降りる，2) 坂を登る/下る，3) 点字ブロックを通過するという路面状態に関する行動

[1] http://opportunity-project.eu/challenge

[2] open door 1, open door 2, close door1, 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

[3] walking forward, walking left, walking right, walking upstairs, walking downstairs, running forward, jumping, sitting, standing, sleeping, elevator up, elevator down

[4] http://www.motionnode.com/

に特化したラベルが提供されている．

4.2.2 ニューラルネットの構造

ディープネットの構造には単純な全結合ネットワークを積み重ねたものや，時間方向を考慮した Recurrent Nueral Networks などさまざまな構造が存在しており，行動認識に対してどの構造が優れているかの検討は進められている [12, 19, 21]．特に，ネットワーク構造に関する比較としては Hammerla らの論文が詳しい [21]．ここでは，ネットワーク構造を畳み込みニューラルネットに固定し，深いネットワーク構造が既存の特徴量と比較してユーザ独立認識に有効な特徴を獲得することを確認する．

具体的には，6 層の畳み込みニューラルネット（CNN）と，最大 24 層で構成される Residual Networks [15] を利用した．CNN では，今回利用するデータセットの 1 つである Opp データセットを利用した研究である [19] を参考に，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)400 個のユニットを持つ全結合層，4) 出力層の合計 6 層のネットワークを利用した．なお，畳み込みサイズ（フィルタ数）は順に 1×5(50)，1×5(40)，1×3(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．最適化には Adam [101] を利用し，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．学習エポック数は最大 300 エポックまでとし，検証データによる早期終了により認識精度が 10 エポック改善しなかった場合には学習を打ち切った．また，ドロップアウト [92] を利用して正則化を行った．

より深い構造をもつニューラルネットとして，He らが提案した Residual 構造 [15] を利用したネットワーク（Res_L）を検証に利用した．Residual 構造とは図 4.1-(a) のように単純な順伝播パスと迂回パスを利用したものである．このような Residual 構造を利用することにより，深いネットワークの学習が可能であることが知られている．図 4.1-(b) に，Residual 構造を利用したネットワークの全体像を示す．ここで N は，Residual 構造を何度繰り返すかを示すものであり，また Convolution および Res Block の右上に記した 1×5 (50) というのは畳み込みサイズとフィルタ数をそれぞれ意味している．N=0, 1, 2, 3 のネットワークを学習し，それぞれの層数 L はそれぞれ L=6, 12, 18, 24 であった．最適化には，CNN と同様 Adam を利用し，学習率は 0.0001 とした．

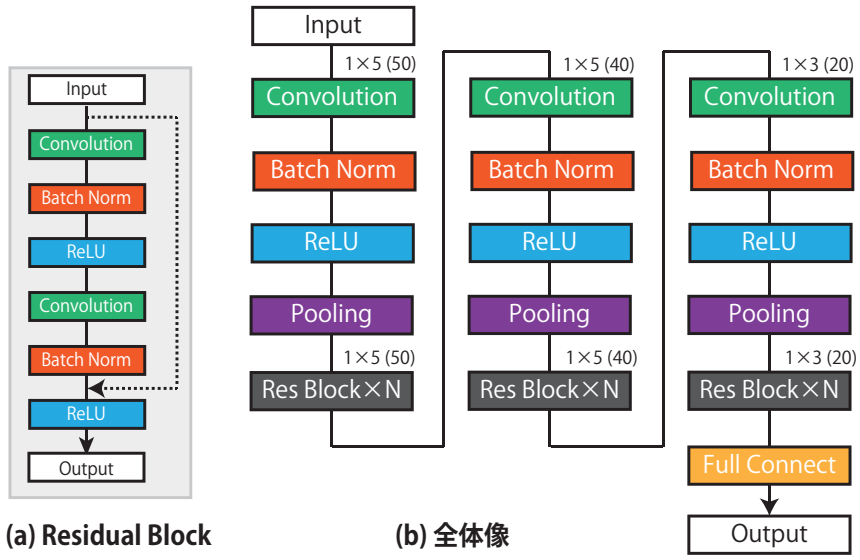


図 4.1: 分析に利用した Residual Networks の概要，(a)Residual Block の詳細，(b)Residual Block を利用したネットワーク構造の全体像．

4.2.3 比較対象

ディープネットの有効性を検証するための比較対象としては，次の 7 つを利用した．

- **Raw** [102, 19]: 先行研究に従い，生の時系列データを入力として利用する K 近傍分類器（KNN）を訓練した．K の値として K =1, 5, 10, 15, 20, 25, および 30 を検証し，検証データに対して最も良い精度となった K=1 の場合の認識精度を報告する．

- **MV** [102, 19]：Raw と同様に，K =1 の KNN を利用した．特徴量には入力各チャネルの平均と分散を利用した．

- **SVM** [102]：特徴量としては MV と同様入力各チャネルの平均と分散を利用し，分類器に Support Vector Machine（SVM）を利用した．カーネルにはガウシアンカーネルを利用し，パラメタ C およびに γ についてはグリッドサーチにより検証し検証データに対して最も良い精度となるパラメタを選択した．

- **Heu** [102]：特徴量として各チャネルの平均と分散に加えて，最大値，最小値，FFT 変換後の直流成分，FFT 変換後のエントロピー，FFT 変換後のエネルギーの 7 種類を利用した．分類器としては KNN を利用した．

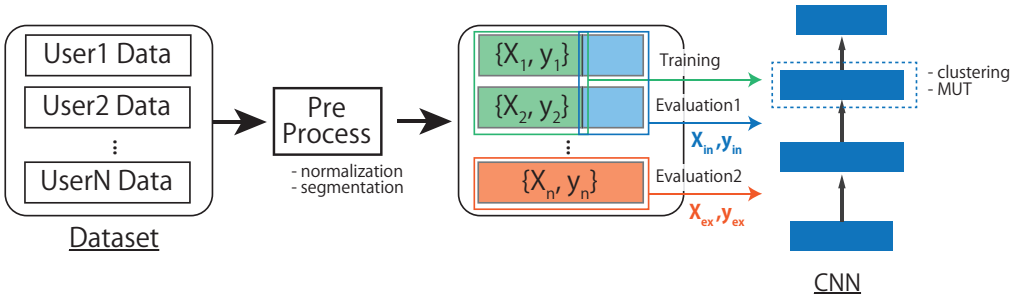


図 4.2: ディープネットの学習および評価の流れ.

- **ECDF** [12, 11]：ECDF（Empirical Cumulative Distribution Function）は，Plötz ら [12] により初めて提案され，Hammerla [11] らにより複数のデータセットでの有効性が検証された手法である．ECDF は，経験累積分布の逆関数により任意の分布の特徴を保存した特徴量である．詳細は，[11] を参照のこと．

- **+MV** CNN が学習した特徴に加えて，MV で利用した素性を加えた分類器を作成した．具体的には，CNN の全結合層 400 ユニットと MV 素性を単純に結合しロジスティック回帰を利用して分類した．

- **+Heu** +MV と同様，CNN の全結合層 400 ユニットと Heu 素性を単純に結合しロジスティック回帰を利用して分類した．

ここで前者 5 つが経験的に利用されてきた素性のみを利用したものであり，後者 2 つは CNN に加えて経験的に利用されてきた素性を利用するものである．これらの 7 つとの比較を通して，1)CNN が学習する表現のユーザ独立認識への適性，2)CNN が学習する表現と経験的に利用されてきた素性の包含関係について検証する．

4.3 評価方法

図 4.2 にディープネットのユーザ独立認識に対する有効性の検証方法の流れを示す．まずディープネットを学習するために，前処理によりサンプルを抽出した．次に，得られたサンプルのうち一部のデータを利用し学習し，残りのデータで認識精度を定量的に検証した．また，学習されたニューラルネットがどのような特徴を獲得しているのかについて，学習された表現空間上でサンプルをクラスタリングすることにより検証した．それぞれの要素について順に説明する．

表 4.1: 分析に利用したデータセットのメタ情報.

Dataset	#Users	#Samples	#In	#Out	Max %	Min %
Opp	4	49,881	113	18	0.62	0.01
Wheel	9	7,016	3	4	0.70	0.04
Hetero	9	870,819	3	7	0.17	0.12
USC	14	172,169	6	12	0.14	0.04

4.3.1 前処理

評価にあたり，まず前処理として一般的によく利用される正規化とスライディングウィンドウ方式によるセグメンテーションを行った．正規化では各チャネルごとに平均 0 分散 1 になるような変換を加えた．正規化後の入力データをスライド幅 w 重なり率 p のスライディングウィンドウ方式により分割した．セグメンテーション後に得られるサンプルは $\mathbf{X_i} = [\mathbf{x_t}, \mathbf{x_{t+1}}, \cdots \mathbf{x_{t+w-1}}]$ のようになる．ここで，$\mathbf{x_t}$ は時刻 t において各センサが観測した値の k 次元のベクトルであり，k は利用したセンサの数であり表 4.1 中#In に対応する．結果として得られたサンプルの数（#Samples），入力次元数（#In），出力次元数（#Out），もっとも多いクラスの割合い（#Max），もっとも少ないクラスの割合（#Min）は表 4.1 のようになった．

4.3.2 ユーザ独立認識性能の評価方法

ユーザ独立認識性能の評価にあたり，一般的に利用される Leave-One-Subject-Out 手続き（LOSO）により評価を行った．LOSO とは，K 人のユーザのうち $K-1$ 人分（訓練ユーザ，図 4.2 中緑色およびに水色）のデータにより学習したモデルを，残りの 1 人のユーザ（テストユーザ，図 2 中橙色）のデータを利用して評価することを K 回繰り返し行う方法である．このようにある特定のテストユーザ以外のデータを利用して学習したモデルに対する評価値を K 回得ることで，未知のユーザに対する認識精度を評価した．

各試行における評価用のデータセットとしては，次の $\{X_{in}, y_{in}\}$ およびに $\{X_{ex}, y_{ex}\}$ の 2 種類を利用した．$\{X_{in}, y_{in}\}$ は評価の対象としているネットワークの訓練に利用したユーザと同一のユーザのデータであり（図 4.2 中の水色部分），$\{X_{ex}, y_{ex}\}$ は学習に利用していないユーザのデータである（図 4.2 中の橙色部分）．この 2 つのデータセットに対する認識精度およびにその違いを確認することで，ネットワークが獲得する表現が新規ユーザに対してどの程度の認識性能を持つのか，またどの程度学

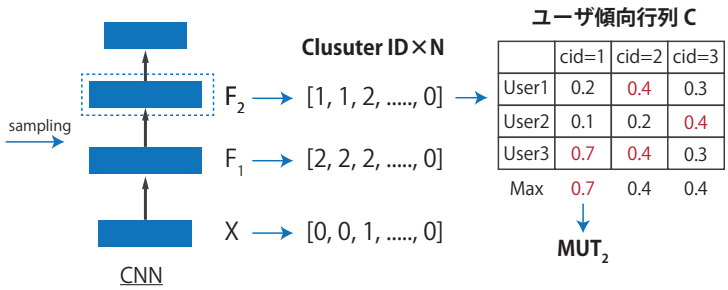


図 4.3: 最大ユーザ傾向の計算方法.

習ユーザに対して過学習するかを確認した.

評価指標としては，正解率およびに平均 F 値を利用した．正解率およびに平均 F 値の定義は次のとおりである.

$$accuracy = \frac{1}{N}\sum_{n=1}^{N} 1(\hat{y_n} = y_n),$$

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

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に予測ラベル，$F^l(\hat{y}, y)$ 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るという点で異なる役割を持つ指標である.

4.3.3 学習された表現のクラスタリングによる評価

認識精度による定量的な評価に加えて，学習された特徴量の可視化を通して学習された特徴量のユーザ依存性を分析した．具体的には，次の手順により計算するユーザ傾向行列 C およびに最大ユーザ傾向 MUT（図 4.3）により評価を行った.

手順 1：入力データ X の特徴量への変換

学習されたモデルに入力データの一部（X）を入力し，特徴量（F_l）へ変換した．ここで，F_l とは，l 番目の層における特徴量である．この際，X に含まれるユーザ数，行動の数が等しくなるようにサンプリングした.

手順 **2**：特徴量 F のクラスタリング

手順 1 で得られた特徴量 F を k 個のクラスタに分割した．クラスタリングした結果得られたクラスタ ID のベクトルと各サンプルのユーザ ID をクロス集計することで，図 4.3 中右上のような行列 C を得た．ただし，ここで行列の要素 c_{ij} はユーザ i がクラスタ j に占める割合であり，クラスタ j のユーザ i への依存度の強さと解釈できる．なお，クラスタリング手法としては k 平均法を利用した．

手順 **3**：最大ユーザ傾向の算出

得られた行列 C を使い，最大ユーザ傾向（Maximum User Tendency，MUT）を下記式に基づき計算した．

$$MUT_l = \max\{c_{ij}, i = 1, 2, \cdots, n, j = 1, 2, \cdots, k\}$$

MUT は k 個のクラスタのうち最もユーザ依存度の高いクラスタにおけるユーザ依存度を意味する指標であり，大きいほど特徴量がユーザに依存していることを意味する．例えば，MUT が 0.8 であれば k 個にクラスタリングした場合に，少なくともいずれか 1 つのクラスタでは特定のユーザが 80%以上を占めていたことを意味する．

4.4 実験結果

4.4.1 CNN と経験的特徴量の比較

表 4.2-(a) に CNN と各比較手法のテストデータ X_{ex} に対する認識精度の比較した結果を示す．また，表 4.2-(b) に検証データ X_{in} に対する認識精度との差を示す．双方の表において，<u>太字＋下線</u> の項目は認識精度がもっとも高かったことを，太字の項目は 2 番目に高かったを意味している．

まずテストデータに関する認識精度に着目すると，F 値およびに正解率の双方で平均すると CNN を利用した手法群（CNN，+MV，+Heu）が顕著に良い精度となっていた．具体的には，F 値，正解率が CNN は 0.627，0.740 であり，経験的に利用されてきた特徴群の中で最も良い精度である Heu の 0.556 およびに 0.665 と比較してそれぞれ約 0.07 ポイント，0.08 ポイントずつ上回っていた．また，テストデータに対する Wilcoxon の符号付き順位和検定を CNN と KNN–ECDF まで行った結果，いずれも p <0.01 で有意差ありとなった．

検証データについても平均するとCNNを利用した手法群の精度が良くなっているものの，データセットごとに見ると例えばWheelではECDF，HeteroではRaw，USCではHeuが最も良くなるなどテストデータに対する認識精度ほど顕著ではなかった．また，経験的に利用されてきた特徴量の中で最も良い精度であったECDFとCNNの2群間でWilcoxonの符号付き順位和検定を行ったところ，F値と正解率についていずれもp値 > 0.05 でありいずれも危険率5%で有意差なしという結果となった．

表 4.2: CNN と既存の特徴量の認識精度における比較.

(a) テストデータ X_{ex} に対する認識精度

	F 値								正解率							
	KNN	MV	SVM	Heu	ECDF	CNN	+MV	+Heu	KNN	MV	SVM	Heu	ECDF	CNN	+MV	+Heu
Opp	0.318	0.468	0.468	0.480	0.463	0.550	**0.557**	**0.559**	0.639	0.698	0.698	0.714	0.700	0.774	**0.776**	**0.776**
Wheel	0.222	0.569	0.569	0.622	0.621	**0.704**	0.702	**0.712**	0.364	0.785	0.785	0.823	0.823	0.879	**0.880**	**0.882**
Hetero	0.527	0.480	0.564	0.513	0.482	0.640	**0.644**	**0.645**	0.521	0.472	0.579	0.506	0.475	0.681	**0.682**	**0.682**
USC	0.439	0.384	0.452	0.480	0.420	**0.570**	0.542	**0.583**	0.450	0.407	0.469	0.494	0.433	**0.582**	0.558	**0.594**
average	0.411	0.512	0.536	0.556	0.532	**0.627**	0.624	**0.636**	0.527	0.626	0.655	0.665	0.642	**0.740**	0.736	**0.744**

(b) 検証データ X_{in} に対する認識精度

	F 値								正解率							
	KNN	MV	SVM	Heu	ECDF	CNN	+MV	+Heu	KNN	MV	SVM	Heu	ECDF	CNN	+MV	+Heu
Opp	0.713	0.833	0.833	0.836	0.845	0.874	**0.875**	**0.876**	0.848	0.915	0.915	0.916	0.923	0.929	0.929	0.929
Wheel	0.234	0.706	0.706	0.752	**0.780**	0.757	0.752	**0.759**	0.381	0.861	0.861	0.886	0.898	**0.899**	0.899	**0.900**
Hetero	**0.739**	0.650	0.682	0.692	0.697	0.697	0.697	**0.699**	**0.753**	0.662	0.713	0.707	0.711	**0.740**	0.738	0.739
USC	0.691	0.644	0.683	**0.777**	0.759	0.755	**0.764**	0.758	0.716	0.680	0.718	**0.803**	**0.787**	0.763	0.771	0.764
average	0.537	0.653	0.670	0.700	0.704	**0.714**	0.713	**0.717**	0.614	0.721	0.743	0.761	0.761	**0.775**	0.775	**0.776**

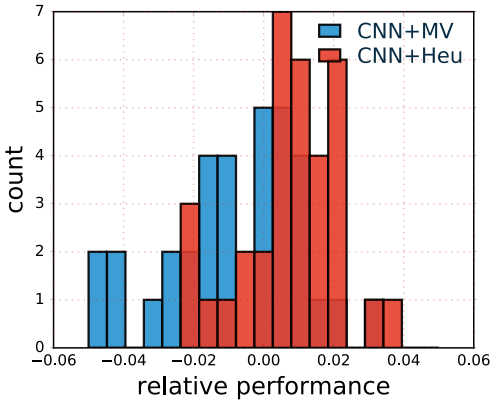
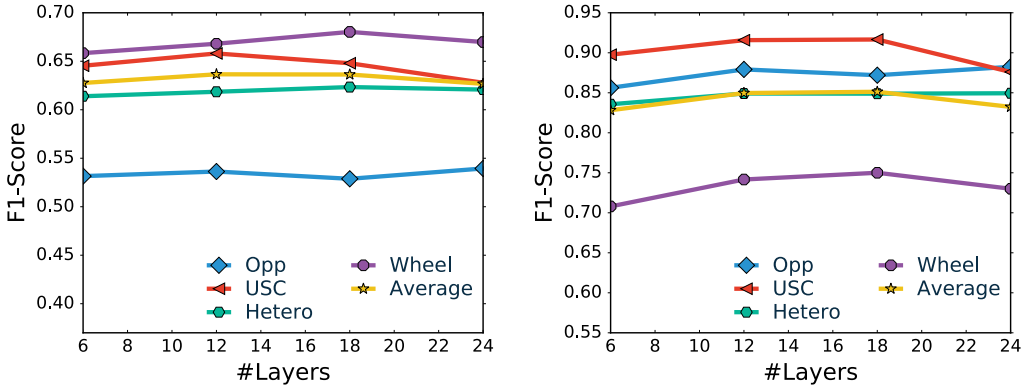


図 4.4: 経験的な特徴量の追加による認識精度の変化度合いに関するヒストグラム．

CNN と CNN に各素性を加えた場合を比較すると，CNN+Heu 素性の結果がテストデータに対してすべてのデータセットでわずかに良い精度となっていた．図 4.4 に CNN に MV 素性およびに Heu 素性を加えた場合の認識精度の変化度合いをヒストグラムとして可視化した結果を示す．具体的には，テストデータに対する F 値について，CNN からの改善度合いを横軸にしたものを描画している．この結果から，MV 素性を入れた場合には認識精度が多くの場合微減（最大 0.05）しており，また平均でも認識精度に低下が見られた．一方，Heu 素性を入れた場合には認識精度が多くの場合微増（最大 0.04）していた．また，全体を平均した改善は微小（0.009 ポイント）であったものの，p 値 < 0.01 で有意差ありとなった．

4.4.2 Residual Networks による分析

より深い構造を持つネットワークを利用した場合の結果として，図 4.5 に Res_L の結果を示す．図 4.5-(a) がテストデータに対する認識精度を，図 4.5-(b) が検証データに対する認識精度をそれぞれ示している．横軸が層の数（6 層，12 層，18 層，24 層で実験）を表しており，認識精度には平均 F 値を利用した結果を示している．結果として，層の数を増やした場合に検証データに対する認識精度は高まる傾向にあるのに対して，テストデータに対する認識精度はほぼ横ばいとなっていた．6 層と 12 層，18 層，24 層のネットワークをそれぞれ Wilcoxson の符号付き順位和検定により比較した結果，検証データに関してはそれぞれ 2.47×10^{-7}，2.47×10^{-7}，0.0005 となりそれぞれ危険率 5%で有意差が見られた一方で，テストデータに関してはそれぞれ 0.112，0.076，0.394 となり有意差はなかった．



(a) テストデータ X_{ex}　　(b) 検証データ X_{in}

図 4.5: $\mathrm{Res}_{\mathbf{L}}$（$L$=6, 12, 18, 24）を利用した場合の認識精度の比較.

4.4.3　行動クラスごとの分析

表 4.3 に，行動クラスごとに認識精度を比較した結果を示す．データセットとしては，USC データセットについての結果を示している．ここで USC データセットについての結果を示すのは，表 4.2 の結果から検証データとテストデータに対する認識精度の差が大きかったことと，行動のラベル名がわかりやすいためである．また，図 4.6 に検証データとテストデータに対する混合行列の差分を可視化した結果を示す．濃い青色ほど検証データに対する値が大きく，赤色が濃いほどテストデータに対する値が大きいことを意味している．例えば，(WF, WF) が濃い青色になっているのは検証データに対しては正しく認識できた WF の認識精度が大きく低下していたことを，(WF, WR) の赤色が濃くなっているのは検証データでは生じていなかった WF を WR と誤認識することが頻出していたことを意味している．

これらの図表から次のような結果が得られた．第 1 に，認識精度の低下度合いには行動クラスごとに大きな開きがあった．例えば，Walking Fowrard はその差分が 0.31 ポイントと大きいのに対して，Running Forward は 0.06 ポイントとその差は軽微であった．開きが小さいものとしては，他に Jumping Up があり，逆に大きいものとしては Walking Right や Walking Downstairs などがあった．

第 2 に，誤りが発生しやすい行動にはパターンが見られた（図 4.6）．例えば，Walking Left や Walking Right，Walking Upstairs などは Walking Forward と誤分類されやすく，また Standing や Sleeping は Sitting と誤分類されやすい傾向にあった．

表 4.3: 行動クラスごとの認識精度.

	検証 F 値	テスト F 値	差分 F 値
Walking Forward (WF)	0.71+-0.01	0.40+-0.12	0.31+-0.12
Walking Left (WL)	0.58+-0.02	0.37+-0.11	0.21+-0.10
Walking Right (WR)	0.63+-0.01	0.37+-0.10	0.26+-0.10
Walking Upstairs (WU)	0.77+-0.02	0.56+-0.20	0.21+-0.22
Walking Downstairs (WD)	0.75+-0.02	0.51+-0.12	0.23+-0.12
Running Forward (Rw)	0.87+-0.01	0.81+-0.09	0.06+-0.10
Jumping Up (JU)	0.87+-0.01	0.78+-0.12	0.09+-0.13
Sitting (Si)	0.86+-0.01	0.64+-0.22	0.21+-0.22
Standing (St)	0.90+-0.01	0.73+-0.18	0.17+-0.18
Sleeping (Sl)	0.93+-0.01	0.77+-0.19	0.16+-0.19
Elevator Up (EU)	0.55+-0.05	0.39+-0.16	0.15+-0.14
Elevator Down (ED)	0.65+-0.02	0.51+-0.15	0.14+-0.15
Average	0.75+-0.02	0.57+-0.15	0.18+-0.15

4.4.4 特徴量のユーザ依存性の定性評価

ここまで認識精度による評価を中心に行ったが，ここでは，特徴量のユーザ依存性の評価を行うために，特徴量のクラスタリングを行い，得られるクラスタのユーザ傾向行列 C を可視化することによりそのユーザ依存性を定性的に評価する．図 4.7 にクラスタ数 k =20 の場合の Opp および Hetero データセットにおけるユーザ傾向行列 C を示す．ここで，ユーザ傾向行列 C の各行はユーザを，各列はクラスタ ID を意味しており，青色が濃いブロックは特定のユーザがそのクラスタ全サンプルのうちに占める割合が高いことを意味する．例えば，Opp に関する結果はクラスタ 17 や 18 のほとんどが S2 ユーザのものであることを，Hetero に関する結果は例えばクラスタ 7 のほとんどがユーザ e であることを示している．クラスタが特定のユーザにかたよるということは利用している特徴量がユーザを識別するような特徴量を持っていることを意味しており，この結果を言い換えると，学習された特徴量は依然としてユーザごとに異なる値をとっていることを示唆している．なお，この結果はデータセットや学習ユーザによらずほぼ同様に生じていた．

図 4.8 に，CNN の各層ごとの MUT を比較したものを示す．横軸がクラスタリングの際に利用したクラスタ数であり，縦軸が MUT の値である．また，色及びにマーカーがどの層の特徴量を利用したかを意味している．なお，Input とは入力 X における MUT を，Conv1–Conv3 はそれぞれ各畳み込み層を，FC が 400 ユニットの全

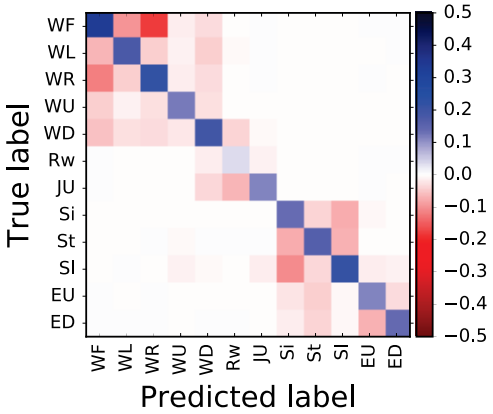


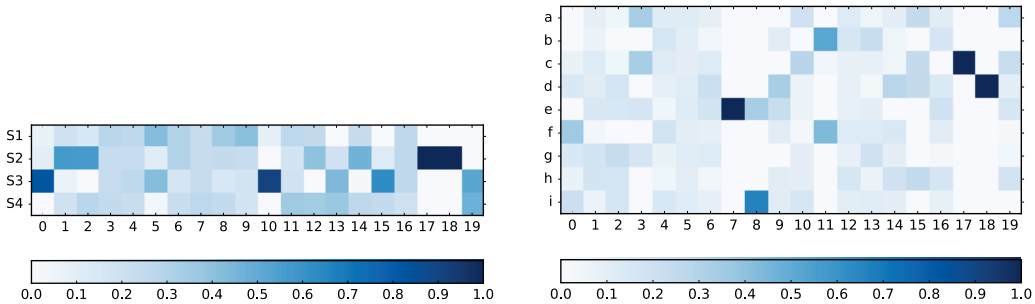
図 4.6: テストデータと検証データでの混同行列の差分の可視化.

結合層をにおける MUT を意味している．Input を除く 4 つにおいてはクラスタリングを行う際の特徴量の値として活性化関数を通した後のものを利用した．また，ここで示した MUT の値は Leave-One-Subject-Out の手続きで得られた n 個のモデルにおける平均値である．

結果として，次のような結果が観察された．1. データセットおよびに層の位置によるものの，多くの場合クラスタ数を 15 にしたあたりで MUT が 0.8 を超えていた．この結果は，多くの場合 80%以上が特定のユーザで占められるようなクラスタが現れていたことを意味している．ただし，USC データセットにおける FC 層のみはユーザ依存度が低い傾向にあった．2. Conv1–Conv3 および に FC 層における MUT を比較すると，浅い層（Conv1 側）ほど MUT が高い傾向にあった．3. Hetero や USC データセットにおいて，ニューラルネットにより処理したあとの MUT の方が Input の MUT より高くなっていた．

4.5 関連研究・考察

これらの結果をまとめると，まず，6 層の畳み込みネットワークの利用により経験的に利用されてきた特徴量より正確なユーザ独立認識が可能になることが示された．表 4.2-(a) が示すように，単純な CNN が経験的な特徴量群（MV，Heu，ECDF）よりいずれのデータセットでも良い精度となっていた．また，CNN が学習した特徴量に Heu 素性を追加することで認識精度は有意に改善したものの，その改善はわずかであった（図 4.4）．この結果は，特に今回利用した経験的な素性については，CNN が十分に代替するような特徴を学習していたことを示唆している．本研究と同じよ



(a) Opportunity (b) Hetero

図 4.7: クラスタ数を 20 とした際の各クラスタのユーザ傾向の可視化．(a)Opportunity, (b)Heterogeneous

うにディープネットの行動認識に対する有効性を検証した研究として，例えば Plötz らは行動認識タスクにおいてディープネットをはじめて利用し，従来の特徴設計パイプラインを制限付きボルツマンマシンや主成分分析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ている [12]．近年の研究で，Jing らは畳み込みニューラルネットが今回も利用した Opp データセットで高い認識精度となることを報告している [19]．また，Ordóñez や Hammerla らは Long Short Term Memory のように時系列性を考慮したネットワーク構造が優れていることを報告している [20, 21]．一方，これらの研究はいずれもユーザ依存認識での検証にとどまっていた．本研究は，これらの研究を背景に，学習データに対して過学習しすぎないことが求められるより難しい問題であるユーザ独立認識においてもディープネットが有効に機能することを示したものである．

このように CNN により獲得される特徴量は経験的に用いられてきた素性と比較して良い認識精度となることが示された一方，Lane らや Reiss らが指摘する新規ユーザに対する認識精度の低下問題 [37, 55] は CNN においても依然として残っていた．例えば，表 4.2 の結果は，Opp データセットや USC データセットでは学習時に存在するユーザに対する認識精度と存在しないユーザに対する認識精度の差はそれぞれ F 値で 0.324，0.185 ポイントと大きなものとなることを示していた．図 4.5 の結果は，18 層などのより深い構造のネットワークを利用することは，学習時に存在するユーザに対する認識精度を改善する一方で存在しないユーザに対する認識精度にはほぼ影響しないか，あるいは低下させる場合もあることを示唆していた．認識精度の開きは特徴的な行動（Jumping Up や Running Forward）については軽微であった（表 4.3）ものの，Walking Forward と Walking Upstairs など，より類似した細か

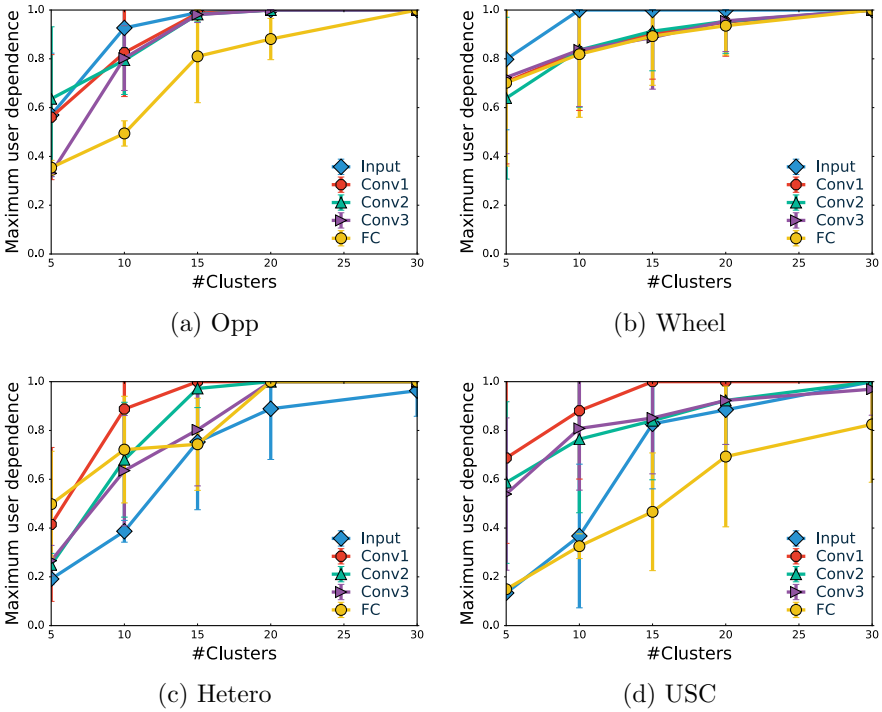


図 4.8: クラスタ数を変化させた際の各層における MUT の比較

な行動の違いを識別する際には甚大であった．これらの結果は，特に精緻な行動の分類を不特定ユーザに対して行うアプリケーションを想定する場合には表現力を適切に選ぶ必要があることを示唆している．

また，図 4.7 に示したように，特徴空間でクラスタリングを行うと特定のユーザに偏ったクラスタが生成された．言い換えれば，学習された特徴空間上における各サンプルの位置はどのユーザであるかと言う情報を含むものである可能性を示唆していた．このような特徴量のユーザ依存性はユーザ独立認識精度を高めることを難しくするだけでなく，プライバシーに関する問題を引き起こす可能性もある．例えば，特徴量がユーザの属性（年齢や性別，あるいは疾病の有無など）の違いにより異なる値を取るようなものであった場合，そのような特徴量を利用することは暗黙的にユーザの属性情報を収集してしまう可能性がある．この状態は，ユーザ側からすれば同意しない情報を取得されるリスクを，アプリケーション提供者としては意図しない個人情報を取得して無用なプライバシーリスクを負う可能性があり望ましくない．プライバシーへの配慮は行動認識を実用的に行う上では重要な要素であり，実際にディープネットを実用的に利用する際には学習されたモデルがどのような情報を持ちうるのかについて慎重に検討する必要があると考える．

ユーザ依存性を低減させるためのより直接的なアプローチとして，関連領域であるドメイン適応やドメイン汎化などの領域で利用されている技術とニューラルネットの結合がありうる．この領域では，ドメイン適応の分野で利用されている Maximum Mean Discrepancy(MMD) [103] を利用した Deep Domain Adaptation [104], Variational Fair Auto-Encoder [105]，2 つの分布の距離を分類器により図る H ダイバージェンス [103] を利用した Domain Adversarial Neural Networks [106] などが提案されている．また，ドメイン汎化の領域ではニューラルネットベースのモデルとして Contractive Auto-Encoder [107] や Multi-Task AutoEncoder [108] が利用，提案されている．これらの手法を発展させていくことが学習される特徴量のユーザ依存性を低減し，ユーザ独立認識精度を高めるうえで重要な将来課題であると考えられる．

第5章　ユーザ敵対的ニューラルネット：敵対的訓練によるユーザ独立な表現学習法

5.1　表現のユーザ独立性

4章で見たように，単純なディープネットの適用によりユーザ独立認識の精度は高まるものの，ディープネットは高い表現力を持つため学習時に存在するユーザに対して過学習しがちであり，また学習される特徴量はユーザに依存したものになっていた．特徴量がユーザ依存になっていると，未知のユーザに対する汎化性能に影響が出ることに加えて，本来計測したかった行動に関する情報以外を収集してしまう可能性がある．例えば，仮に行動認識のための特徴量としてユーザに依存する歩き方などの利用を許容すると，その背後にあるユーザの属性情報（年齢，性別，疾病の有無など）を暗黙的に収集してしまう可能性がある．この状態は，ユーザ側からすれば同意しない情報を取得されるリスクを，アプリケーション提供者としては意図しない個人情報の取得により無用なプライバシーリスクを負う可能性があり望ましくない．ユーザ独立な特徴を利用すること，言い換えればユーザ（あるいはその属性）に依存するような特徴を利用しないように制約をかけることで，このようなプライバシー問題を低減することができると考えられる．

本章では，学習時に敵対的ユーザ分類器を考慮することで学習ユーザに対して過学習しにくくなるように制限する提案手法について述べる．ここで表現のユーザ独立性とは，形式的には表現を表す確率変数 R とユーザを表す確率変数 U の独立性のことを意味する．ユーザ独立ではない，すなわちユーザ依存であるとは特徴量の値がユーザごとに異なることを意味する．例えば，あるユーザの歩き方のみに強い値を取るような特徴量や，ユーザにより値が異なるような特徴量はユーザ依存な特徴量である．

ここで，ユーザ独立性を高めることは究極的には行動の予測性能とトレードオフ

になりうることは重要である．例えば，単純にユーザ独立性を高めるのでであれば完全にランダムにデータをマッピングすれば良いが，そのような写像は明らかにデータの持つ重要な性質を取り除いてしまう．すなわち，プライバシーに配慮しながら適切に行動認識を行うためにはユーザ独立性を高めることと行動に関する予測精度を高めるという 2 つの問題を考える必要がある．

本章ではユーザ独立な特徴を獲得する問題を，特徴量 R のもつ行動 Y に関する情報を最大化しながら，ユーザ変数 U に関する情報を最小化する問題として定式化する．そのうえで，上記の最適化問題を解く方法として敵対的訓練（adversarial training）を利用したユーザ敵対的ニューラルネット（User-Adversarial Neural Networks）を提案する．敵対的訓練によるユーザ独立な特徴獲得のアイディアは，特徴量 R からユーザ変数 U を予想しようとする敵対的ユーザ分類器 f_u を騙すように特徴抽出器 $f_f : X \to R$ を学習させることである．すなわち，提案手法では識別器 f_u が R からどの程度 U を予測できるかをユーザ独立性の指標へと利用する．敵対的訓練を利用したディープネットの拡張により，ディープネットによる高い識別能力を活用しながらユーザ独立な特徴量を獲得できることが期待される．

本章では，提案手法によりユーザ独立な表現が獲得できることを慣性センサを利用したパブリックな行動認識データセット 2 つ（3 つの分類タスク）を利用して検証する．具体的には，行動認識精度とユーザ認識精度の 2 つの観点を用いて，行動認識でよく利用される畳み込みニューラルネット（CNN）と比較して提案手法が行動認識精度を維持しながらユーザ独立な表現を獲得することができることを定量的に検証する．また，学習された特徴空間でのクラスタリングにより得られるクラスタのユーザ独立性を視覚化し，表現のユーザ独立性が改善していることを定性的に確認する．これらの検証を通じて，提案手法により既存の手法と比較してよりユーザ独立な特徴が獲得できることを示す．

本章の主要な貢献は次の通りである．

- ユーザ独立な特徴を獲得する問題を行動に関する認識精度を最大化しながらユーザに関する認識精度を最小化する問題として定式化した．その上で，最適化問題を解くニューラルネットのモデルを提案した （5.3 節）．また，情報理論的独立性の指標である相互情報量と提案手法の関係を整理することで，提案手法の妥当性について論じた．

- 実データによる検証により，提案手法が単純なディープネットと比較してよりユーザ独立な特徴を確認できることを示した （5.4 節）．この結果は，敵対的

訓練によるユーザ独立性の指標の利用が実際にユーザ独立性を高めることに貢献することを示したものである．

- 実験により得られた結果から，ユーザ独立性およびにユーザ汎化性能を高めるための課題，プライバシーへの配慮をより良く行うための課題をまとめた（5.5 節）．

以降まず，本章で利用する敵対的訓練について関連研究を交えて説明する．その後，敵対的訓練をプライバシーへの配慮に応用する提案手法について説明し，実験により提案手法の有用性を検証する．

5.2 関連研究

敵対的訓練によるユーザ依存性最小化のアイディアは，特徴量 R からユーザ変数 U を予想しようとする識別器 u を騙すように，特徴抽出器 $f : X \to R$ を学習させることである．すなわち，識別器 u がどの程度 R から U を予想できるかを R のユーザ依存度の指標へと利用する．

筆者が知るかぎり，敵対的訓練アプローチは Schmidhuber らにより初めて利用された [109]．より近年の研究で，Goodfellow らは敵対的訓練による方法を画像生成タスクに利用し [69]，さまざまな研究による改良がなされている [71]．提案手法により近い研究として，Ganin がドメイン適応の文脈で提案した Domain Adversarial Neural Networks（DANN）[110] と Edwards が表現から機微情報を取り除くための手法として開発した Adversarial Learned Fair Representations（ALFR）がある [111]．これらはある特徴量 R が何らかの情報 S（前者はドメイン，後者は機微情報）に依存しないように学習するという点では提案手法と類似しているが，学習される特徴量のユーザ依存性を低下させることを目的とする本手法とは異なる目的で開発されているものである．

また，これらの研究ではドメインや機微情報 S として 2 値を想定しているのに対して，本稿が取り扱うユーザ独立な表現を獲得する問題では U は一般に多値となるという点でも異なっている．U が多値になることが引き起こす問題として，1. DANN や ALFR は H ダイバージェンスという 2 つの分布（$P_{s=1}(R)$ と $P_{s=0}(R)$）の距離を測る指標に理論的に立脚しているが，H ダイバージェンスは多値を直接扱えないため多値にした場合の理論的妥当性が不明である，2. 取り除きたい情報 S が 2 値の場合と比較して，敵対的分類器と特徴抽出器のバランスをとることが難しくなる，と

いった問題がある．これらの問題に対して本稿では，1. 独立性の情報理論的な指標である相互情報量と提案手法の関係を整理，2. バランスを取るための技術的対策の提案を行い，その有用性を検証する．

5.3 ユーザ敵対的ニューラルネット

5.3.1 定式化

前述したとおり，ユーザ独立かつ行動認識に利用できる特徴量を獲得するためには，特徴量 R が Y に関する情報を持ちながら U に関する情報を持たないようにする必要がある．この問題は，次のような 2 つの損失関数の同時最適化問題として定式化することができる．

$$\mathrm{L} = \mathrm{L}_y(X, Y) + \lambda \mathrm{L}_u(R, U) \tag{5.1}$$

ここで，L_y および L_u は特徴量が持つラベル Y に関する情報の大きさおよびユーザ U に関する情報の小ささを表す損失関数である．また，λ は相対的な重みである．以降，この最適化問題を順伝播型ニューラルネットで解く提案手法，ユーザ敵対的ニューラルネットについて説明する．

5.3.2 UANN の構造

図 5.1 にユーザ敵対的ニューラルネットの構造を示す．参考までに，通常の順伝播型ニューラルネットを横に並べた．図中矢印は，順伝播パスおよび逆伝播パスを意味している．提案手法は，通常のニューラルネットが持つ 1. 特徴抽出器 $f_f : X \to R$, 2. 分類器 $f_y : R \to Y$ に加えて，3. 敵対的ユーザ分類器 $f_u : R \to U$ の 3 つのニューラルネットにより構成される．ここで，f はそれぞれ任意の構造を持つニューラルネットであり，それぞれ $\theta_f, \theta_y, \theta_u$ のパラメタを持つ．提案手法では，これらの 3 つのニューラルネットを利用して L_y および L_u を定量化する．

5.3.3 $\mathrm{L}_y(X, Y)$ の定量化

L_y は，通常のニューラルネットと同様の方法で求める．すなわち，X と Y の間で定義される交差エントロピーを最小化することで求める．交差エントロピーによる

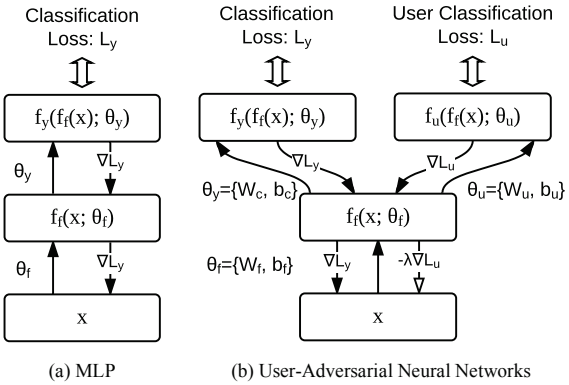


図 5.1: ユーザ敵対的ニューラルネットの構造と一般的な MLP の比較.

損失関数は，定式的には次の式のようになる．

$$\mathrm{L}_y = -\frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M} y_{nm} \log f_y(f_f(x_n)) \tag{5.2}$$

ここで，N は学習サンプル数，M は分類したいクラスの数，y_{nm} はサンプル n が m 番目のクラスであるかを表す $\{0,1\}$ の値である．特徴抽出器および行動分類器は確率変数 X から Y が予想しやすいようにする．すなわち式 5.2 を最小化するようにパラメタ θ_f および θ_y を修正する．

5.3.4 $\mathrm{L}_u(R,U)$ の定量化

L_u の定量化には敵対的訓練を利用する．具体的にはまず，変数 R から U を予想しようとするような敵対的分類器 f_u を考える．これは，f_u を任意の構造のニューラルネットとすると，次の損失関数を最大化することにより実現できる．

$$\mathrm{L}_u = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K} u_{nk} \log f_u(f_f(x_n)) \tag{5.3}$$

ここで，K はユーザの数，u_{nk} はサンプル n が k 番目のユーザであるかを表す $\{0,1\}$ の値である．ユーザ分類器は，確率変数 R からユーザ U を予測するように関数 f_u を学習する．これは，形式的には特徴抽出器 f_f のパラメタである θ_f を固定した上で，式 5.3 を最大化するようにパラメタ θ_u を更新することと対応する．

逆に，特徴抽出器は，敵対的分類器が確率変数 R からユーザ U を学習できないようにする，これは形式的には，敵対的ユーザ分類器 f_u のパラメタ θ_u を固定した上

で式 5.3 を最小化するようにパラメタ θ_f を更新することで実現できる．また，概念的には特徴 R から U を予想するような敵対者を想定した上で，その仮想的な敵対者が情報を抽出できないように f_f を修正することに対応する．

5.3.5 最適化

まとめると，提案手法は 1. 特徴抽出器 f_f，2. 行動分類器 f_y，3. 敵対的ユーザ分類器 f_u の 3 つの要素から構成され，最終的な最適化問題は式 5.4 のようになる．

$$\min_{\theta_f,\theta_y}\max_{\theta_u}[\mathrm{L}_y(X,Y;\theta_f,\theta_y)+\lambda\,\mathrm{L}_u(R,U;\theta_f,\theta_u)] \tag{5.4}$$

全体の最適化問題は明らかに凸でなく，単純な最適化で解くのは難しい．ここでは，[106] や [111] に従い確率的最急降下法（SGD）による交互最適化を行った．すなわち，1. L_u を最大化するように θ_u を更新する，2. L を最小化するように θ_f および θ_y を更新するという手続きを交互に繰り返し行った．

敵対的訓練の成否の鍵をにぎる要素の 1 つとして，L に関する最小化と最大化のバランスがあることが知られている [69, 106, 111]．ここでバランスとは，提案手法の場合敵対的ユーザ分類器が L_u を最大化する能力と特徴抽出器が敵対的ユーザ分類器を騙す能力の均衡のことを意味する．このバランスが崩れると，勾配が十分に伝わらなかったり，おかしな方向に進んだりするおそれがある．特に，今回の場合，敵対的分類器が行うのは多クラス分類となるため，2 値分類である類似研究より敵対的分類器の認識能力が低くなりがちであると予想される．実際，実験的にも初めから L_u を最小化するように特徴抽出器を学習させると過度にユーザ独立にする能力が強くなり行動の予測精度が上がらない現象が観測された．

この問題を解消するために，λ をはじめは小さな値からはじめエポックが進むごとに徐々に大きくするアニーリングを行った．パラメタ λ のアニーリングを行うことで，ユーザ独立性の観点において特徴抽出器が特に初期フェーズで過剰に強くなるのを抑えることができる．スケジューリングとしては，今回は単純に線形に増やす方法を利用した．本稿の実験で利用した具体的なスケジューリング式は以下のとおりである．ここで，t はエポック数，$\alpha+1$ はアニーリングを開始するエポック，β は

Algorithm 1 Optimization of the proposed model

Require: Samples $S = \{(x_n, y_n, u_n)\}_{i=1}^{N}$
Require: hyper parameters λ
Ensure: neural network $\{\theta_f, \theta y, \theta u\}$
  $\theta_f, \theta y, \theta u \leftarrow$ initialize()
  t $\leftarrow 1$
  while training() **do**
    $\lambda_t \leftarrow Eq.5.5$
    $\mathrm{L}_u \leftarrow Eq.5.3\ \forall(x_n, u_n) \in S$
    $\theta_u \leftarrow \theta_u + \frac{\delta L_u}{\delta \theta_u}$ {Update User-Adversarial classifier s.t maximize L_y}
    $\mathrm{L}_y \leftarrow Eq.5.2\ \forall(x_n, y_n) \in S$
    $\mathrm{L}'_u \leftarrow Eq.5.3\ \forall(x_n, u_n) \in S$
    $\mathrm{L} = L_y + L'_u$
    $\theta_f \leftarrow \theta_f - \frac{\delta L}{\delta \theta_f}$
    $\theta_y \leftarrow \theta_y - \frac{\delta L_y}{\delta \theta_y}$
    t $\leftarrow$ t + 1
  end while

アニーリングを終了するエポック，λ は最終的なパラメタの設定値を意味している．

$$\lambda(t) = \begin{cases} 0.0 & (t \leq \alpha) \\ (t-\alpha) \cdot \frac{\lambda}{(\beta-\alpha)} & (\alpha \leq t < \beta) \\ \lambda & (\beta \leq t) \end{cases} \tag{5.5}$$

以上をまとめた提案手法の擬似コードは Algorithm1 の通りである．

5.3.6 相互情報量最小化との関係

情報理論的には，確率変数 R が確率変数 U に関する情報を持つかは 2 つの確率変数の独立性により表される．2 つの確率変数の独立性は，次に示す相互情報量により定義することができる．

$$I(U;R) = \sum_{r \in R} \sum_{u \in U} P(u,r) \log \frac{P(u,r)}{P(u)P(r)}$$

$I(U;R)$ は 0 以上の値であり，0 の時に R と U は独立である，すなわち R は U に関する情報を含まないことを意味する．

ここで，L_u と $I(U;R)$ の間には次のような対応関係がある．詳細は B.1 を参照の

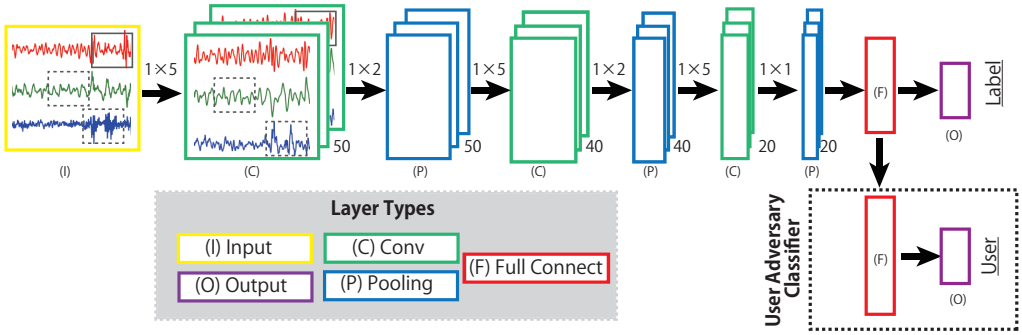


図 5.2: 実験に利用した提案手法のネットワーク構造.

こと.

$$I(U;R) = N \cdot \mathrm{L}_u - C$$

ただし，C は定数である．すなわち，θ_f を固定して L_u を最大化することは θ_u を相互情報量 $I(U;R)$ を最大化するように更新することを，また θ_u を固定して L_u を最小化することは θ_r を相互情報量 $I(U;R)$ を最小化するように更新することを意味している．

5.4 実験

5.4.1 データセット

Opporunity Activity Recognition

データセットとして，最もポピュラーなデータセットの 1 つである Opportunity データセットを利用した．Opportunity データセットは，朝食をシミュレートしたシナリオにおける日常行動を 4 人分（S1–S4）記録したデータセットである．多数の種類の活動者の身体，物体，および環境に設置されたセンサにより行動を計測しており，合計で 72 のセンサ，10 のモダリティを持つデータである（詳しくは，Sagha らの論文の図 1 [98] を参照されたい）．結果として，各時刻において 113 の実数値（時刻情報は除く）が記録されている．本稿では，提供されている 3 つのうち Gestures（以降 Opp-G）と Locomotion（以降 Opp-L）による検証を行った．Opp-G は，18 ク

ラス[1]．のより細かな行動を分類するタスクであり，Opp-L は 5 種類のより大局的なユーザの状態を分類するタスクである．分類問題を解くにあたり，一般に利用されるスライディングウィンドウ方式（スライド幅 30，重なり率 0.5）により学習サンプルを生成し，結果として 4 ユーザ合計で約 50,000 のサンプルを得た．

USC-HAD

USC-HAD は，公開されている行動認識データセットの中では比較的大規模な男女 7 名ずつ合計 14 名のデータを含む点が特徴である [112]．行動としては，前に歩く，右に曲がる，左に曲がるなどの一般的な生活における最も基本的な 12 種類の行動が記録されている[2]．データの計測は MotionNode[3] により行われており，6 次元の系列データである．Opportuniry データセットと同様，スライディングウィンドウ方式により学習サンプルを生成し，結果として 14 ユーザ合計で約 127,000 のサンプルを得た．

5.4.2 実験設定

図 5.2 に提案手法を利用した User-Adversarial CNN の具体的なネットワーク構造を示す．敵対的ユーザ分類器を除くと，ネットワーク構造は単純な CNN と同様となる．層の数やフィルタ数などのハイパーパラメタは Opportunity データセットに対して CNN を利用した研究である [19] を参考に設計した．ここで，図中 $1 \times K$ は畳み込みサイズ（あるいはプーリングサイズ）を，各層の右下に位置する数値は特徴マップ数を意味しているものである．活性化関数としては出力層にソフトマックス関数を利用したほかは，ReLU を利用した．また，各畳み込み層および全結合層のあとにはドロップアウトを行い，過学習を抑制した．最適化には Adam [101] を利用し 150 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．

敵対的ユーザ分類器 f_u としては，図中全結合層における特徴量を入力とする MLP_{800} を利用した．ここで，MLP_{800} とは 800 個の中間ユニット数もつ 3 層ニューラルネッ

[1] open door 1, open door 2, close door1, 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

[2] walking forward, walking left, walking right, walking upstairs, walking downstairs, running forward, jumping, sitting, standing, sleeping, elevator up, elevator down

[3] http://www.motionnode.com/

トを意味している．L_u と L_y を重み付けるハイパーパラメタ λ は他に言及がない限り 0.1 に設定した．また，アニーリングのパラメタ α および β は開始直後と終了直前の 15 エポックずつは固定した λ で学習されるように，それぞれ 15，135 とした．

提案手法の有効性およびパラメタの重要性を検証のため次の 4 つの手法を比較した．1) CNN：提案手法において λ=0.0 とした場合に対応．2) $\mathrm{UANN_{LR}}$：敵対的ユーザ分類器として，$\mathrm{MLP_{800}}$ の代わりにロジスティック回帰を利用したものに対応．3) $\mathrm{UANN_{MLP}}$：提案手法においてパラメタ λ のアニーリングを行わない場合に対応．4)$\mathrm{UANN_{MLPA}}$：敵対的ユーザ分類器として $\mathrm{MLP_{800}}$ を利用し，かつアニーリングを行った場合．これらの手法を比較することで，単純な CNN と比較して提案手法群がよりユーザ独立な特徴量を獲得できること，敵対的ユーザ分類器の表現力およびアニーリングの影響を検証する．

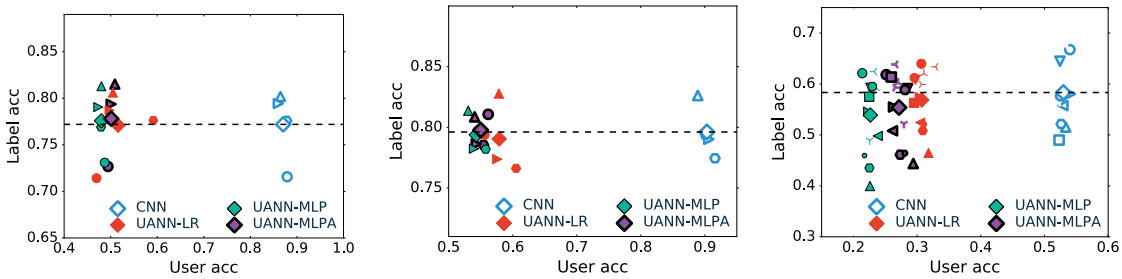
5.4.3 評価方法

各手法の評価は次のような Leave-One-Subject-Out 手続き（LOSO）により行った．LOSO は新規ユーザに対する認識精度を検証するためによく利用される方法であり，ここでは 1. K 人のうちから 1 人テストユーザを i を選択する，2. ユーザ i を除く $K-1$ 人のデータの一部を訓練データとして利用しモデルを訓練する（このとき，80%を訓練データとし残り 20%は検証データとして利用した），3. ユーザ i のデータ（以降テストデータ）を利用して評価を行う，という手続きを K 回に渡り繰り返し行う．このような手続きでは，ユーザ i のデータは各ステップのモデルにおいて訓練に全く利用されないので，学習されたモデルの新規ユーザに対する認識性能評価することができる．また，テストデータによる評価に加えて，検証データに対する認識精度とテストデータに対する認識精度の差を比較し，学習されたモデルのユーザに関する過学習度合いを評価した．

モデルの評価指標として，行動分類能力に関する指標と表現のユーザ独立性に関する指標の双方を利用して評価を行う．行動分類能力に関する指標としては，正解率および平均 F 値を利用した．正解率および平均 F 値の定義は次のとおりである．

$$accuracy = \frac{1}{N}\sum_{n=1}^{N} 1(\hat{y_n} = y_n),$$

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



(a) Opp-G (b) Opp-L (c) USC-HAD

図 5.3: 行動認識精度とユーザ独立性のバランスの比較．各マーカはテストユーザ，各色は手法の違いを表している．また，ダイヤモンドマーカは全ユーザの平均値を表す．

ここで，N は全サンプル数，L は行動クラスの集合，y_n と $\hat{y_n}$ 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るという点で異なる役割を持つ指標である．

ユーザ独立性に関する定量的な指標としては，ユーザの認識精度を利用した．具体的には，モデルを学習後にその特徴量 R からユーザを分類するような分類器 γ を改めて学習させ，その認識精度を計測した．このとき認識精度が高い，すなわち特徴量 R からユーザ U が予測できる場合には独立性が低く，逆に予想できない場合には独立性が高いことを意味する．認識精度は用いる分類器 γ に強く影響されると考えられるため，ここでは γ としてロジスティック回帰（LR），隠れユニット数 50 の 3 層ニューラルネット（MLP_{50}），および隠れユニット数 800 の 3 層ニューラルネット（MLP_{800}）の 3 つを利用した．

ユーザ独立性に関して，定量的な評価に加えてクラスタリングによる定性評価を行った．具体的には，学習されたモデルの特徴量 R をクラスタリングし，クラスタリング結果から算出したユーザ傾向行列 C を可視化した．ここで，C の要素 c_{ij} はクラスタ j においてユーザ i のサンプルが占める割合を示しており，0.0 から 1.0 までの値を取る．C を可視化することで，特定のユーザに偏ったクラスタが発生しているかどうか，すなわち学習された特徴量が強いユーザ依存性を持つかどうかを可視化した．

表 5.1: 行動認識精度に関する比較.

(a) テストユーザに対する認識精度

	平均 F 値				正解率			
	CNN	$\mathrm{UANN_{LR}}$	$\mathrm{UANN_{MLP}}$	$\mathrm{UANN_{MLPA}}$	CNN	$\mathrm{UANN_{LR}}$	$\mathrm{UANN_{MLP}}$	$\mathrm{UANN_{MLPA}}$
Opp-G	**0.557**	0.547	0.551	**0.566**	0.772	0.771	**0.776**	**0.778**
Opp-L	**0.794**	0.786	0.781	**0.804**	**0.796**	0.790	0.794	**0.798**
USC	**0.573**	**0.548**	0.501	0.523	**0.583**	**0.568**	0.539	0.554
average	**0.660**	0.643	0.624	**0.644**	**0.735**	**0.724**	0.713	0.721

(b) ユーザに関する過学習度合い

	平均 F 値				正解率			
	CNN	$\mathrm{UANN_{LR}}$	$\mathrm{UANN_{MLP}}$	$\mathrm{UANN_{MLPA}}$	CNN	$\mathrm{UANN_{LR}}$	$\mathrm{UANN_{MLP}}$	$\mathrm{UANN_{MLPA}}$
Opp-G	-0.321	-0.333	**-0.319**	**-0.303**	-0.159	-0.161	**-0.151**	**-0.147**
Opp-L	**-0.159**	-0.165	-0.171	**-0.148**	**-0.145**	-0.151	-0.147	**-0.143**
USC	-0.187	-0.149	**-0.104**	**-0.120**	-0.185	-0.139	**-0.094**	**-0.109**
average	-0.234	-0.231	**-0.216**	**-0.205**	-0.161	-0.152	**-0.135**	**-0.136**

5.4.4 実験結果

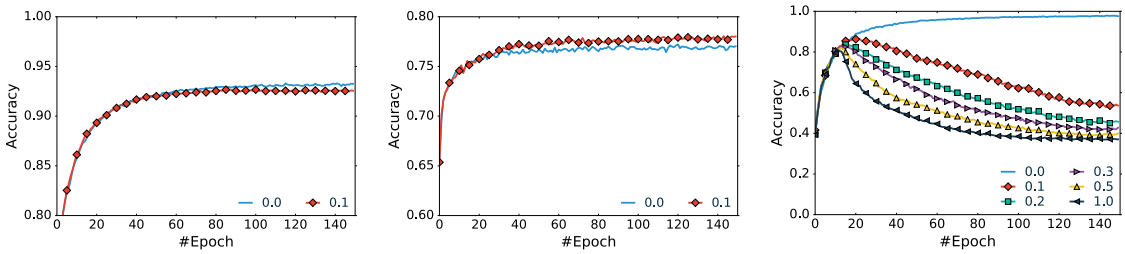
行動認識精度とユーザ独立性のバランス

図 5.3 に，各データセットにおける行動認識精度およびユーザ認識精度を比較した結果を示す．縦軸が行動認識精度を，横軸は分類器としてロジスティック回帰を利用した場合のユーザ認識精度を表しており，左上に行くほどユーザ独立かつ行動の認識精度が高く，右下に行くほど行動の認識精度が低くユーザ独立性も低いことを意味している．各マーカは異なるユーザを表しており，各色がそれぞれ異なる手法に対応している．なお，図中の点線は CNN の行動認識精度を示したものである．

結果として，まずユーザ認識精度に関して，CNN と比較して提案手法群の認識精度が著しく低下していることがわかる．行動認識精度はデータセットにより結果が異なっているものの，Opp-G および Opp-L では提案手法群の認識精度は CNN と同程度か僅かに上回っており，また USC-HAD データセットについても認識精度の低下はユーザ認識精度の改善具合と比較すると軽微であった．この結果は，敵対的訓練を利用した提案手法により行動認識精度とユーザ独立性のより良いトレードオフを実現できることを示唆している．

ユーザ汎化性能に関する結果

表 5.1 に，各データセットでの行動認識精度の比較結果を示す．表 5.1-(a) はテストユーザに対する認識精度を，また表 5.1-(b) は参考までにユーザに関する過学習度



(a) 行動認識精度（検証データ） (b) 行動認識精度（テストデータ） (c) ユーザ認識精度

図 5.4: 行動認識精度およびユーザ認識精度のエポックごとの変化.

合い（テストデータと検証データに対する認識精度の差）を示している．太字＋下線の項目は認識精度がもっとも高かったことを，太字の項目は 2 番目に高かったことを意味している．

結果として，次のような結果が得られた．まず，CNN と $\mathrm{UANN_{MLPA}}$ を比較すると，新規ユーザに対する認識精度については平均 F 値および正解率双方で Opp-G および Opp-L については $\mathrm{UANN_{MLPA}}$ が優れており，USC については通常の CNN が優れていた．また，ユーザに関する過学習度合いでは USC を含むすべての場合で $\mathrm{UANN_{MLPA}}$ が優れていた．この結果は，提案手法が利用した敵対的訓練は学習時に存在するユーザに対する認識精度を低下させるものの，新規ユーザに対する認識精度は僅かに改善するかあるいは低下する場合でもその低下度合いは学習時に存在するユーザに対するものと比較すると軽微であったことを意味している．

次に，$\mathrm{UANN_{LR}}$ と $\mathrm{UANN_{MLPA}}$ を比較すると，USC データに関するテストユーザに関する認識精度を除くと $\mathrm{UANN_{MLPA}}$ が上回る結果となっていた．また，Wilcoxson の検定を行うとテストユーザに関する認識精度では平均 F 値および正解率それぞれ 0.055，0.061 の p 値となり 5%危険率で有意差は見られなかったが，過学習度合いでは平均 F 値および正解率それぞれ 0.0005，0.0002 の p 値となり有意差が見られた．この結果は，敵対的ユーザ分類器としてより適切な分類器を利用することでユーザに関する過学習を抑制できる可能性を示唆している．最後に $\mathrm{UANN_{MLP}}$ と $\mathrm{UANN_{MLPA}}$ を比較すると，ユーザに関する過学習度合いについては同程度であった一方で，新規ユーザに対する認識精度では $\mathrm{UANN_{MLPA}}$ がすべての場合で上回っていた．この結果は，パラメタのスケジューリングを行うことで過学習を同程度に抑えながらテストユーザに対する認識精度を適切に高めることができることを示唆している．

図 5.4-(a) および (b) に，検証データとテストデータに対する CNN（青線）と $\mathrm{UANN_{MLPA}}$（赤線）の認識精度の推移を比較した結果を示す．ここで，横軸がエ

表 5.2: 異なるユーザ分類器を利用した場合のユーザ認識精度の比較.

(a) Opp-G

	LR	MLP_{50}	MLP_{800}
None	0.870	0.963	0.983
LR	0.516	0.955	0.981
MLP_{800}	**0.501**	**0.774**	**0.950**
Rand	0.333	0.333	0.333

(b) Opp-L

	LR	MLP_{50}	MLP_{800}
None	0.903	0.966	0.980
LR	0.579	0.953	0.977
MLP_{800}	**0.550**	**0.822**	**0.957**
Rand	0.333	0.333	0.333

(c) USC-HAD

	LR	MLP_{50}	MLP_{800}
None	0.530	0.649	0.770
LR	0.307	0.493	0.636
MLP_{800}	**0.272**	**0.377**	**0.521**
Rand	0.078	0.078	0.078

ポック数，縦軸が正解率である。また，図中の値は Opp-G における 4 回の試行の平均をとったものである．なお，今回の実験では α=15 としたため，15 エポック目までは CNN と $UANN_{MLPA}$ は同じ損失関数を最適化している．結果として，1) 約 20 エポック目までは両者はほぼ同じ正解率になっていること，2) 20 エポックを超えたあたりから $UANN_{MLPA}$ が安定的に良い精度になっていること，3) 検証データに関しては 100 エポックを超えたあたりで $UANN_{MLPA}$ が飽和しているのに対して CNN では更に学習が進んでいること，の 3 点が確認できた．これらの結果は，単純な CNN はユーザに対して過剰に適合する可能性があること，また提案手法により学習ユーザに対する過剰適合を抑制できていることを示唆している．

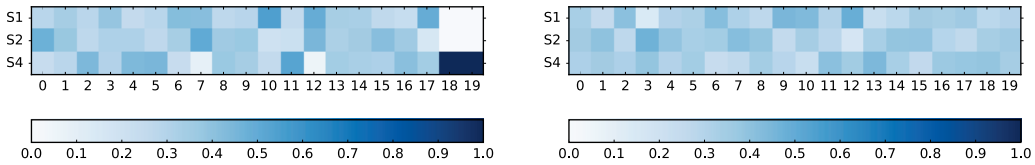
図 5.4-(c) に，検証データに対するユーザ分類器の正解率のごとの推移を示す．横軸はエポック数，縦軸は正解率であり，各色が異なる λ に対応している．結果，c-1) CNN（λ=0.0）の場合最終的に 99%程度でユーザ識別が可能になる一方，$\lambda \geq 0.1$ の場合にはユーザ分類器の精度を抑制することができていること，c-2) λ を大きくするほどユーザ分類器の精度は早く低下するが，最終的な認識精度は飽和すること，c-3) 正解率の低下はおおよそ 20 エポック目から始めており，図 5.4 においてテストデータに対する認識精度の改善が始まったタイミングとほぼ一致していること，などが確認された．

ユーザ独立性に関する結果

表 5.2 に，異なるユーザ分類器を利用した場合の認識精度を比較した結果を示す．各行はそれぞれ学習時に利用した敵対的ユーザ分類器 f_u を，各列は評価に利用した分類器 γ の種類を示している．f_u=None は敵対的訓練を行わない場合であり，CNN に対応するものである．また，完全にランダムに予測した場合の予測精度 (Rand) を参考値として示した．特徴量が完全にユーザ独立な場合どのような分類器を利用してもランダム予測となるため，Rand は論理的な限界値を示すものである．

結果としてまず，評価用分類器 γ によらず MLP や LR を利用した敵対的訓練を行う場合に，行わない場合と比較してユーザ識別性能が低下していることがわかる．この結果は，提案手法により単純な CNN を利用する場合と比較してユーザ独立性の高い表現を学習できていることを示すものである．一方で提案手法の課題として，γ として f_u より強い表現力を利用する場合，敵対的分類器を利用した場合でもユーザ認識精度が低下しにくい傾向が見られた．例えば，Opp-G について，f_u にロジスティック回帰（LR）を，γ に同様の LR を利用した場合には認識精度は 0.579 ポイントと敵対的訓練を利用しない場合（0.903）と比較して 0.30 ポイント以上認識精度を抑制できていたのに対して，γ に MLP_{50} を利用した場合には 0.953 ポイントとなっており，敵対的訓練を利用しない場合（0.966 ポイント）とほとんど変わらない値であった．また，γ として $\mathrm{MLP}_{8}00$ を利用した場合の結果は各データセットでそれぞれ 0.950，0.957，0.521 であり，ランダムな予測と比べて著しく高かった．

最後に，ユーザ独立性の定性的な評価として図 5.5 に，CNN および $\mathrm{UANN}_{\mathrm{MLPA}}$ で学習された特徴量のクラスタリングによって得られたユーザ傾向行列 C を可視化した結果を示す．図中青が濃い部分がクラスタに占める特定のユーザの割合が高いことを，白に近いほど低いことを意味している．また，クラスタリング手法としては k 平均法を利用した．結果として，提案手法によりユーザ傾向が低いクラスタができていることが視覚的に確認された．例えば，CNN の場合クラスタ 18 や 19 で S4 の割合が高くなっているのに対して，提案手法では同様の傾向を持つクラスタは存在しなかった．



(a) CNN　　(b) UANN (λ=0.1)

図 5.5: クラスタ数を 20 とした際の Opp-G における各クラスタのユーザ傾向の可視化．a: CNN，b: UANN (λ=0.1)．

5.5　考察

5.5.1　ユーザ独立性について

実験の結果は，提案手法が行動の認識精度を可能な限り保った上でユーザ独立な特徴量を獲得できることを示唆していた．図 5.3 の結果は，敵対的訓練を行わない通常の CNN が学習した特徴量と比較して提案手法が学習した特徴量は，行動に関して同程度の予測性能を持つ一方で，ユーザを予測しにくい特徴量となっていることを示している．また，図 5.5 に示した特徴量のユーザ傾向の可視化結果は，CNN が獲得した特徴量によるクラスタの一部や過度に特定のユーザが固まる傾向にあったのにたいして，提案手法が学習した特徴量はすべてのクラスタで特定のユーザが固まるような現象はみられなかった．これらの結果は，敵対的ユーザ分類器を学習時に考慮する提案手法により，CNN が学習する特徴量のユーザ独立性を高めることができることを意味している．

一方，ユーザ独立性に関する限界として，表 5.2 に示したように提案手法により得られた表現はランダム予測と比較すると著しく高いユーザ依存性を持っており，完全にユーザ独立な表現を学習しているわけではないという点がある．言い換えれば，本稿での検証により得られた結果は単純な CNN に対する比較優位性を示すものであり，絶対的に十分なユーザ独立性が得られていることを示すものではない点には注意されたい．提案手法の枠組みの中でよりユーザ独立性の高い表現を学習するためには，適切な分類器やハイパーパラメタ λ の選択，敵対的分類器を接続する層の選択，複数の層における敵対的分類器の利用などの方法が考えられる．これらの項目について検証を加えることで，手法の改良を行う必要がある．

5.5.2 ユーザ汎化性能について

ユーザ独立な特徴量を獲得する動機の1つであったユーザ汎化性能については，まず提案手法を利用することでCNNと比較してユーザ過学習度合いを抑制できることが示された（表5.1）．この結果は，特徴抽出器の出力のユーザ独立性を高めることにより，行動分類器のユーザ汎用性を高めるという提案手法のアイディアの妥当性を示すものであると考えられる．一方で，平均すると過学習度合いは3つのデータセット平均で0.136ポイント（F値では0.205ポイント）と依然として大きい値となっていた．この原因としてまず，5.1節に記述したように提案手法によるユーザ独立性が不十分であったことが考えられる．もう1つの可能性として，ある特徴量が与えられたときの行動の真の確率分布 $P(Y|U)$ がユーザによって異なるケースが考えられる．これは，例えばある人が歩くときに生じる観測パターンと他の人が早歩きするときの観測パターンが類似している場合などでが考えられる．このような場合，特徴量がユーザ独立になったとしても予測された結果が正しいかどうかはユーザによってことなることになる．提案手法の限界として，提案手法は特徴量 R のユーザ独立性を高めるためのものであり，ある特徴量が発生した時にどの行動が生じたかの確率分布 $P(Y|R)$ がユーザごとに異なることに伴う認識精度の劣化を考慮することは出来ない．後者の問題を解決するためには，ユーザ適応技術など別の方法を適用する必要があると考えられる．

また，提案手法によるもう1つの課題として，新規ユーザに対する絶対的な認識精度がデータセットによっては劣化する場合が有ることが確認された．具体的には，表5.1が示すように，Opp-LおよびOpp-Gについては提案手法によりユーザ汎化性能が改善する一方で，USCデータセットについては汎化性能が低下していた．

この原因として，次の2つが考えられる．第1に，モデルの持つ表現力が十分でなかった可能性がある．式5.1に示したように，提案手法では通常のCNNと比較してより複雑な損失関数を最適化するように学習している．その結果として，モデルの持つ表現力が不足していた可能性が考えられる．紙面の都合上割愛したものの，実際USCデータセットについては訓練データに対する認識精度（正解率）がCNNの0.776に対して $\mathrm{UANN}_{\mathrm{MLPA}}$ では0.666と著しく低下しており，Opportunityデータセットを利用した2つの分類タスクについては同様の傾向は見られなかった．また，この差はユーザ汎化性能における差（CNNが0.583に対して提案手法が0.554）の差より著しく大きかった．より多くのパラメタを持つネットワーク構造を利用した場合にユーザ汎化性能が改善するかの検証は今後の課題である．

第 2 に，ユーザ独立な特徴量を学習することに伴う問題として，テストユーザと類似したユーザが学習時に存在する場合にはユーザ独立な特徴量を獲得するよりもその類似ユーザに対して適合した特徴を利用することが認識精度改善に寄与する可能性である．この観点から考えると，学習に利用するユーザ数が増えれば増えるほど新規ユーザと類似した特徴を持つユーザが学習データに含まれる可能性は高まる．一方，学習ユーザが増えれば増えるほど学習が難しくなり，結果として認識精度の改善も限定的である可能性がある．例えば，Ogawa らは学習に利用するユーザ数を増やした際のユーザ汎化性能の変化を検証し，その精度改善が 50 人程度で飽和した報告している [50]．また，今回はデータの都合上 LOSO という新規ユーザが 1 人の場合における検証を行ったが，新規ユーザが増えれば増えるほど学習ユーザに類似しているユーザが含まれる可能性は低下するため，提案手法の有効性が増すと考えられる．より大規模なユーザ数のデータセットによる検証により，どの程度のユーザ数を持つときに提案手法の有効性が強く現れるかの検証が必要と考えられる．

5.5.3 プライバシー保護について

提案手法により得られる特徴量のユーザ独立性は，1 章にも述べたようにプライバシー保護の観点からも重要となりうる．一方で，一般にセンサデータは長時間に渡って収集されることから，敵対者は複数のウィンドウを予測に利用することでより正確に予測を行える可能性があることには注意が必要である．例えば特徴量からユーザを予測する本稿で扱った例では，図 B.1 に示したように複数のウィンドウの結果を平均化することにより著しく高い精度でユーザを予測することができるようになる（詳細は付録 B.2 を参照のこと.）．特徴量が完全に独立になれば複数ウィンドウを利用することは全く意味を成さないため，この問題は特徴量の独立性を高めることにより原理的には解決可能であるものの，本稿で検証したような 1 つのウィンドウに対する特徴量 r からの予測可能性だけでなく，複数のウィンドウを利用した場合の予測性能を考慮した形に手法を拡張することでより優れたプライバシー保護が行える可能性がある．

プライバシー保護の観点からみたときのもう 1 つの課題として，より大規模なデータ，特にユーザ数および含まれる属性が増えた場合での検証が必要である．プライバシー保護の観点からは，ユーザ独自の特徴を利用しないということよりむしろユーザに関する機微情報（例えば病気の有無など）に影響された特徴量にならないことが重要である．ユーザに関する精緻な属性情報を持ったデータベースを利用するこ

とでプライバシー保護の観点からの考察を深めることが必要である．

ユーザ数が増えた際の提案手法の問題点として，敵対的ユーザ分類器の学習が難しくなることがある．例えば，今回，最大でユーザ分類器としては 13 人の場合の分類を行ったが，1,000 人のユーザがいる場合に提案手法における敵対的ユーザ分類器は 1,000 クラスの分類を適切に行うように学習する必要がある．このような複雑な分類問題を行うことは技術的に非常に難しい，あるいは学習に相当の時間がかかると考えられる．対応策として，プライバシー保護の領域で利用されている k-匿名性 [113, 114] や l-多様性 [115]，t-近接性 [116] といった概念が参考になる可能性がある．例えば，1,000 クラス分類を諦め k クラス分類器を複数学習させる方法や，その際に機微情報に関する情報がばらつくように k 人を選定する方法などが考えられる．これらの検証を行うことで，より大規模なユーザが存在する場合にも適用可能なように手法の改良が必要と考える．

第6章　半教師あり蒸留による深層行動認識モデルのユーザ適応

6.1　ユーザ適応

5 章では，ユーザ独立認識精度を高めるためのアプローチとして，敵対的訓練を利用することで特徴量のユーザ独立性を高める手法としてユーザ敵対的ニューラルネットを提案した．ディープネット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．

本章ではこの問題の解決に向けて，ディープネット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 $\{(X_i, y_i)\}(\forall i \in 1 \cdots m)$ 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 $\{(X', y')\}$ 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 6.1）．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 [53]．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，ディープネット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．

本章では，蒸留 [94] 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 によりなめらか

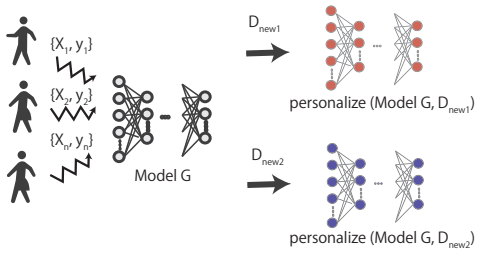


図 6.1: ユーザ適応の概念図.

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ィアの 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 つ目のアイディアは，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 [117] 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 SVD をベースにした手法 [118] および KL 正則化 [117] 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．

本章の主要な貢献は次のとおりである．

- ディープネット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．

- ディープネット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（SSD）を提案した．本手法により，ディープネット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．

本章の残りの構成は次の通りである．6.2節では，まず行動認識でのディープネットの研究動向について概観し，ディープネット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．6.3節では，提案手法について説明する．6.4節では，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に実験結果を説明し，6.5節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章の貢献およびに将来課題について述べる．最後に6.5節で本論文をまとめる．

6.2 関連研究

6.2.1 行動認識でのディープネットの利用

近年，さまざまな研究においてディープネットの行動認識における有効性が検証され，ディープネット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．行動認識においてディープネット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．ディープネットなどの表現学習技術によりデータから直接に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ディープネットの重要性の理由の1つとなっている．

Plötz らは行動認識にディープネットをはじめて利用し，従来の特徴設計パイプラインを制限付きボルツマンマシンや主成分分析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 [12]．また，Volmmer らはスパース符号化を利用したニューラルネットの有効性を報告した [84]．近年の研究で，Jing らは畳み込みニューラルネット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 [19]．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 [20, 21]．これらのように，ディープネット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．

6.2.2 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ディープネットの利用は活発に行われているものの，ディープネット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，ディープネット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ディープネット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 代が歩いている時と 65 歳以上の人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 [37]．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 [48]．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上手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 [53]．Ioffe らはパラメ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 [119]．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 [55]．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，大量のパラメタを持つディープネットをアンサンブル学習させるのは難しいなどの理由からディープネット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 [120] 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 [121, 122]．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 [118]．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 [123]．Siniscalchi らは，活性化関数にパラメ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 [124]．これらのように適応させるパラメ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．

6.2.3 蒸留によるユーザ適応

提案手法は，蒸留 [94] 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で同等の認識精度を得られたり [94]，より望ましい構造を持つニューラルネットに変換することができたり [95]，あるいは特定の条件下で半教師あり学習に有効である [125] 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無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[126] 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，本章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 [117]．また，上述した SVD による重み W を低ランク近似する研究 [118] 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 正則化よりよい制約となると考えられる．

6.3 半教師あり蒸留によるユーザ適応

6.3.1 アルゴリズム

提案手法である半教師あり蒸留（Semi-Supervised Distillation， 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に 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 $\{X^m, y^m\}$ $(\forall m \in 1 \ldots M)$ を利用して不特定行動者モデルを学習する．ここで，X^m およびに y^m はそれぞれユーザ m 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．

$$\mathrm{L}_\theta = \sum_{m=1}^{M} \frac{1}{N^m} \sum_{n=1}^{N^m} \mathrm{L}(y_n^m, \sigma(f_\theta^t(x_n^m))) + \Omega(||f_\theta||), \tag{6.1}$$

θ は任意の構造を持つニューラルネットのパラメタであり，式 6.1 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_n^m はユーザ m の n 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，N^m はユーザ m 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 6.1 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う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用にパラメタ 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 $\{X'_{sup}, y'_{sup}\}$，ラベル無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．

$$
\begin{aligned}
\mathrm{L}_{\theta'} = & \sum_{x,y\in\{X'_{sup},y'_{sup}\}} (1-\lambda)\,\mathrm{L}(y,\sigma(f_{\theta'}(x))) \\
& + \sum_{x\in\{X'_{sup}\}} \lambda\,\mathrm{L}(s,\sigma(f_{\theta'}(x))) \\
& + \sum_{x'\in\{X'_{unsup}\}} \lambda\,\mathrm{L}(s,\sigma(f_{\theta'}(x'))) \\
= & \sum_{x,y\in\{X'_{sup},y'_{sup}\}} (1-\lambda)\,\mathrm{L}(y,\sigma(f_{\theta'}(x))) \\
& + \sum_{x\in\{X'\}} \lambda\,\mathrm{L}(s,\sigma(f_{\theta'}(x))) \\
= & (1-\lambda) \sum_{x,y\in\{X'_{sup},y'_{sup}\}} \mathrm{L}_{hard} + \lambda \sum_{x\in\{X'\}} \mathrm{L}_{soft}
\end{aligned}
\tag{6.2}
$$

ここで，θ' は任意の構造を持つニューラルネットのパラメタであり，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 [94] 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用データセット 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．

6.3.2 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 [117, 118] 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．

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

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，$T \to 0$ 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 [127] 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[126] 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 [128, 129, 130]．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上手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．

6.4 実験

6.4.1 実験設定

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）[98, 99] 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 [19]，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．

ニューラルネットの構造

ニューラルネットの構造としては，不特定行動者モデルおよびに適応後モデルの双方で畳み込みニューラルネットを利用した．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 [19] 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 1×5(50)，1×5(40)，1×3(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．

Baselines

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を学習させた．

[1] http://opportunity-project.eu/challenge

[2] open door 1, open door 2, close door1, 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

- **Fine Tuning (FT, FT-T, FT-D)**: 不特定行動者モデルを式6.1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6.3.2の$\lambda = 0$とした場合に対応する．また，適応後モデル学習再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に全結合層を再学習する場合(FT-D)の3種類を検討した．

- **SVD**[118]: 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．全結合層およびに出力層の重み行列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章では$k = \min(m, n)$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た学習させた．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)** [117]: 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 [101]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ドロップアウト ($p = 0.5$) [92]を利用して学習を行った．他に言及がない限り，提案手法およびにその関係手法のハイパーパラメタとしては$T = 10$およびに$\lambda = 0.8$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ビティに関する分析を行った．

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 $\{X'_{sup}, y'_{sup}\}$ およびに $\{X'_{unsup}\}$ 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に平均F値により評価した．正解率およびに平均F値は次のとおりである．

$$accuracy = \frac{1}{N}\sum_{n=1}^{N} 1(\hat{y_n} = y_n),$$

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

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．

6.4.2 実験結果

提案手法の有用性の検証にあたり，本章では大別して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 およびに 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6.1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）．<u>太字＋下線</u>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，ave_{124} は参考までに不特定行動者モデルの認識精度が極端

表 6.1: 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	0.16	0.15	0.17	0.15	0.19	**0.21**
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**

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.

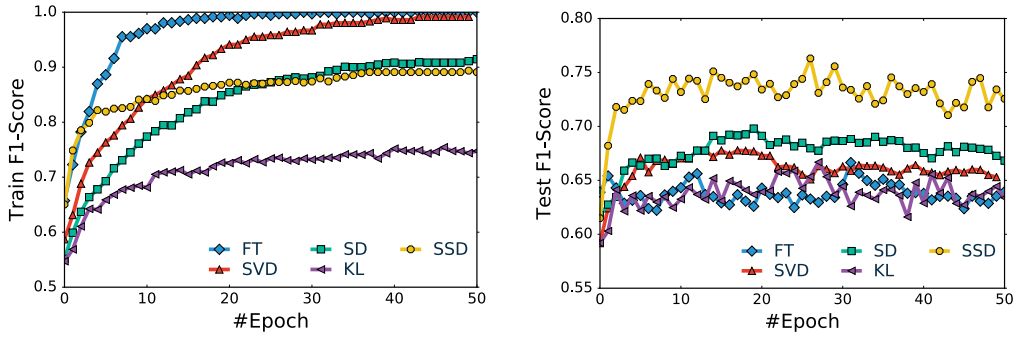
結果として，次のような結果が得られた．1. 提案手法は，$N_{sup} = 5$ の場合 $N_{sup} = 20$ の場合双方で正解率およびに平均 F 値の双方で認識精度に改善が見られた．具体的には，$N_{sup} = 5$ の場合に正解率（平均 F 値）で平均 0.03 (0.07) ポイント，$N_{sup} = 20$

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に平均 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 6.1 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しうる．そのため，表 6.1 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ことではなく，正解率およびに平均 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は注意されたい．

次に，図 6.2 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に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．

6.3 に SSD を行う前後の混合行列および に 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



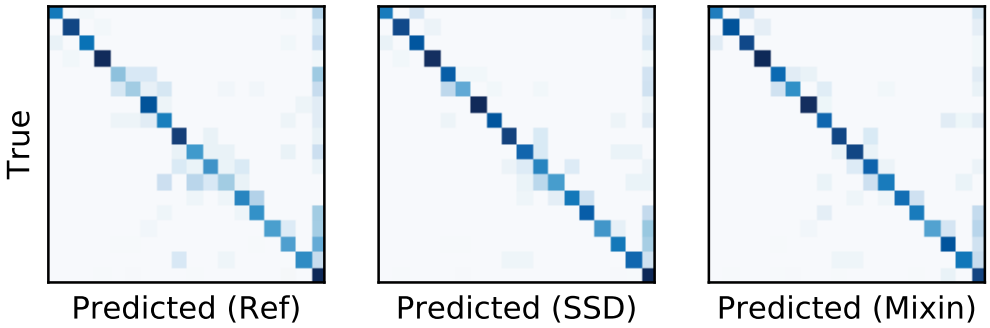
(a) 訓練データに対する認識精度　　(b) テストデータに対する認識精度

図 6.2: 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）

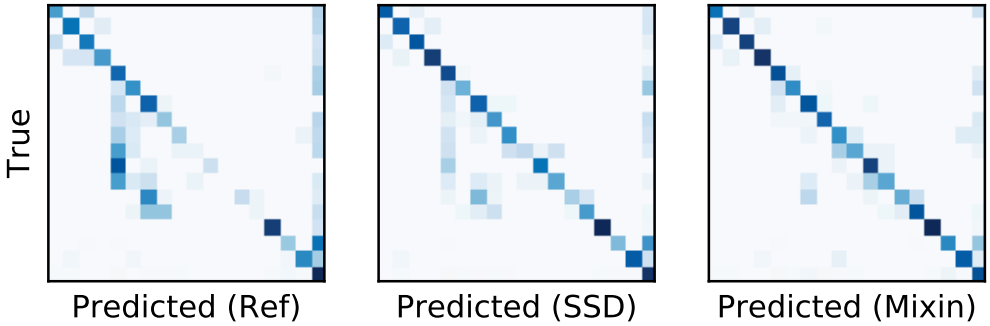
ついては例えば誤って 5 番目のクラスと予測してしまうようなエラーが Mixin と比較して SSD で頻出していることがわかる．この傾向は Ref モデルにおいても同様に起こっており，SSD においては適応の際に元のモデルとして利用した Ref のエラーを引きずるという問題が発生していることを示している．

また，図 6.4 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，図中ところ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 6.5 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 の場合が比較的良い精度となっていた．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



(a) S1

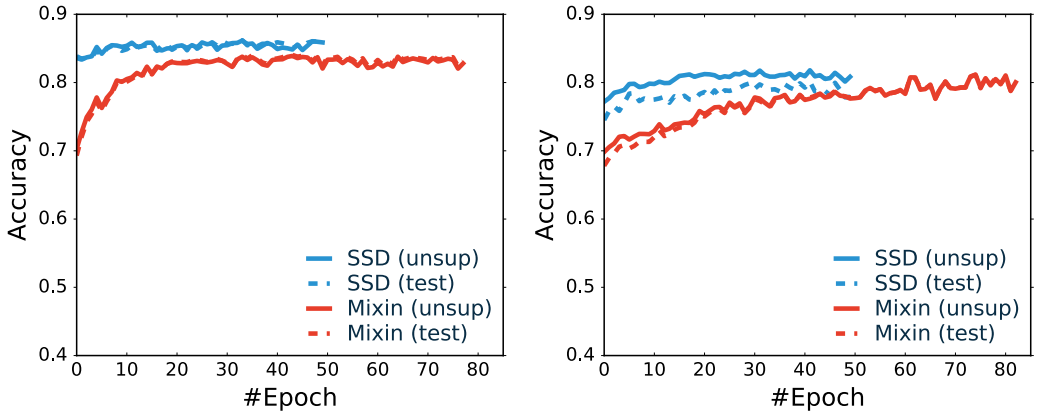


(b) S3

図 6.3: $N_{sup} = 20$ 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．

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 [117] とも整合するものである．

図 6.6 に，提案手法である SSD と SVD を組み合わせた場合の結果を示す．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，



(a) S1　　(b) S3

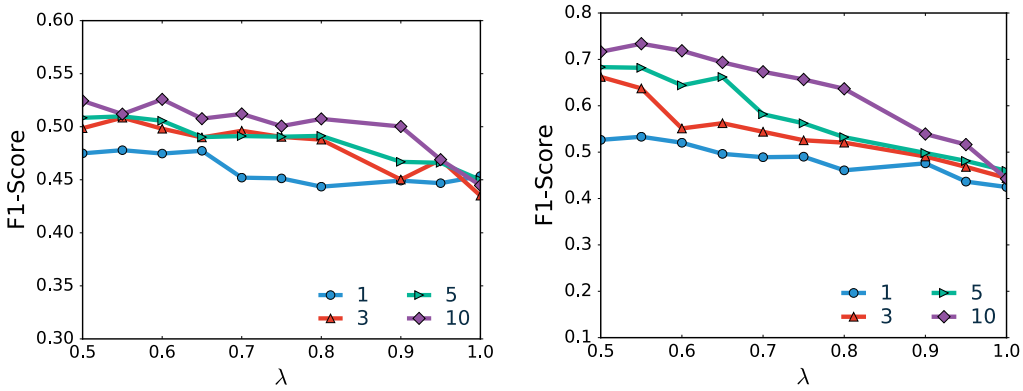
図 6.4: 学習に利用した教師なしデータ X_{unsup} とテストデータ X_{test} に対する学習の進行の比較.

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 6.7 にラベル有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.

6.5　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 6.1 の結果から，ハイパーパラメタおよびに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 6.5 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 6.5 の結果は，ほとんどの場合 T を大きくすることで認識



(a) $N_{sup} = 5$　　(b) $N_{sup} = 20$

図 6.5: T およびに λ に関するセンシティビティの分析.

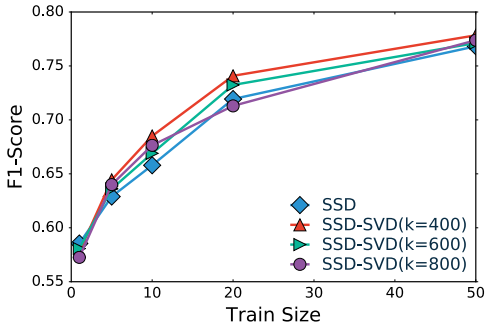


図 6.6: SVD と SSD を組み合わせた場合の認識精度.

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 6.1 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有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 6.7）．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上手く活用することで，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 6.3 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くことを示していた．提案手法が利用している蒸留による正則化は

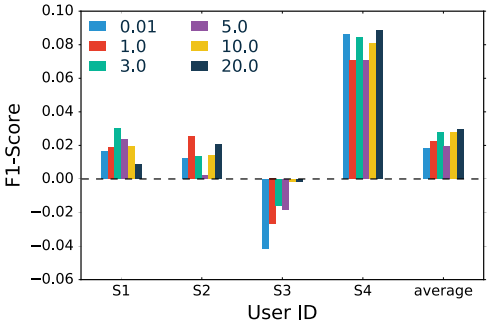


図 6.7: $N_{sup} = 0$ の時の各 T での精度改善度合いの比較.

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 のを徐々に変化さ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ティビティの検証を行ったが，図 6.5 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，ユーザ的応用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用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マ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 6.6 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る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よく学習する手法郡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章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 [123]，ユーザごとに活性化関数の形を変化させる方法 [124]，隠れユニットの出力の大きさを補正する方法 [131, 132] 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 の計算およびに 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を順伝播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．ディープネットのユーザ適応を行う場合に最も計算負荷が大きいのは SGD において誤差を逆伝播させる部分であり，例えばニューラルネットの圧縮技術 [133, 134] と組み合わせるなどの改良の余地がある．

本章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 [135] 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，良い影響を与えると考えられる平均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例えばN_{sup}を均等に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サンプル数用意したり，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に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した選択を行う必要がある．

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場合に，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[55]．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．

第7章　考察

1章でも述べたように，本研究はディープネットに代表される表現学習をウェアラブルセンサによる人間行動の計測技術を推進する上での重要な要素技術であると捉え，ウェアラブルセンサによる人間行動の計測のための表現学習技術の提案を目的とするものである．この目的に対して，本論文ではまず1. ディープネットによる深い表現が認識精度を高めることを確認した上で，ウェアラブルセンサによる行動認識に対して表現学習を適用する上で重要となる2. プライバシーへ配慮した表現の学習およびに3. 教師データの少ないユーザに対して有効な表現の学習という論点に対応した手法の提案を行った．本章では，これらの各論点に対する4章から6章の研究結果をまとめ，また既存研究や事例との関連をふまえて本研究の貢献およびに課題についてまとめる．

7.1　4章から6章の結果まとめ

まず，ディープネットによる深い表現が認識精度を高めるかどうかを検証するという目的に対しては，第4章で約110万サンプル4つのデータセットに対してディープネットの1種である畳み込みネットワークおよびにResidual Networksを適用した．結果として，ディープネットにより学習される表現はユーザ独立認識の設定で従来利用されてきた経験的特徴量と比較して有意に高い認識精度を達成することを示した (表4.2)．この結果は，本論文が5章6章でウェアラブルセンサによる行動認識のための表現学習技術の開発を動機づける本論文の土台となる結果である．

次に，プライバシーへの配慮という目的に対して，4章で既存手法であるディープネットを行動認識データセットに対して適用た結果得られた表現を分析した結果，学習された特徴量は本来獲得したかったわけではないユーザに関する識別性を持つような特徴量となることを確認した (図4.7，図4.8)．4章でも述べたように，この結果が直接的にプライバシーへの重大な問題を引き起こすかについては大規模なデータを利用したさらなる分析が必要であるものの，少なくともディープネットのように認識精度が高いアルゴリズムを利用する場合に意図しない情報を収集してし

まうリスクを示唆する結果であると考えられる．このように表現が意図しない情報を持ってしまう可能性を低減するための手法として，5 章では学習される特徴量がユーザに関する情報を持たなくなるような表現を学習する手法の提案を行った．また，公開された行動認識データセットを利用した実験により提案手法によりユーザに関する情報を持ちにくいような表現を確認できることを示した (図 5.3，図 5.5)．

第 3 に，データが少数のユーザに対して有効な表現を学習するための方法として，不特定多数のユーザにより訓練された汎用的な表現を少数のデータで微調整するユーザ適応手法を 6 章で提案した．また，公開された行動認識データセットを利用した実験により，提案手法によりラベル付きデータが各クラス 5–20 程度と少ない場合であっても高い認識精度を達成することを示した (表 6.1).

これらの知見をまとめると，本論文ではディープネットによる表現により認識精度が従来法と比べて高まること，提案手法により望まない情報を取得する可能性を低減するという意味でプライバシーに配慮するように表現学習を拡張できること，また学習データが少ないユーザの行動を認識する必要がある場合でも他ユーザで学習された表現を修正するユーザ適応技術によって従来より高い認識精度が達成できること示した．これらにより，ウェアラブルセンサによる行動認識を行う上で重要な認識精度およびにプライバシーの双方の観点から，ディープネットによる表現学習が有効になりうることを示した．

7.2 本論文の貢献と課題

7.2.1 正確な行動認識に向けた貢献と課題

次に，以上の研究結果について，既存の行動認識研究における知見をふまえて整理することで本論文で新たに得られた知見およびに本論文の学術的貢献について整理する．まず認識精度の観点から考えると，本研究はディープネットをウェアラブルセンサによる行動認識に適用する試みの一貫として位置づけられる．本研究と同様にディープネットをウェアラブルセンサによる行動認識に対して応用しようとする試みは 2011 年ごろから進められてきた．例えば Plötz らは行動認識タスクにおいてディープネットをはじめて利用し，従来の特徴設計パイプラインを制限付きボルツマンマシンや主成分分析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ている [12]．近年の研究で，Jing らは畳み込みニューラルネットが今回も利用した Opp データセットで高い認識精度となることを報告している [19]．また，Ordóñez や

Hammerla らは Long Short Term Memory のように時系列性を考慮したネットワーク構造が優れていることを報告している [20, 21].

一方，行動認識においてディープネットを実用的に利用するためには，従来の研究が検証を重ねてきたような学習データがリッチに存在するユーザに対する認識精度の観点だけでなく，学習データがほとんど存在しないユーザに対して有効であることが求められる．本研究は，従来研究が示してきた学習データがリッチに存在するユーザに対する認識精度を高める上でディープネットが有効であるという知見を背景に，学習データが殆どないユーザに対する認識精度に関する検証や認識精度を高める工夫を行うものとして整理される．すなわち，2 章で説明した行動認識の類型に照らし合わせれば，従来の研究がユーザ依存認識におけるディープネットの行動認識に対する有効性を示してきたのに対して，ユーザ独立認識やユーザ適応の問題設定で有効性を検証することでより現実的かつ包括的にその有効性を検証したものである．このような包括的な有効性の検討は，行動認識を利用したシステム研究を行っている研究者やアプリケーション開発者がディープネットを有効なツールであるとみなすことにつながり，結果としてアプリケーションの創発に貢献していくものと考えられる．

一方で，ディープネットにより認識精度を高める表現を獲得するという観点を考えたときに，本論文にはいくつかの限界が存在している．ここでは,「どのように表現を獲得するか」という How の観点と,「どのような表現を獲得するか」という What の観点に分けて述べる．

まず,「どのように表現を学習するか」という観点から見たときに，本研究の内容は教師あり学習を利用することを前提にした議論に限定されている．すなわち，本論文での内容は基本的にディープネットを利用して高い認識精度を達成する上で教師あり学習が優れているということを前提においており，例えば教師なし学習の問題設定を考えたときに，6 章で提案したようなユーザ適応手法がそのまま有効であるかは自明ではない．逆に言えば，このような教師なしデータの活用した良い表現の学習手法の開発がウェアラブルセンサによる行動認識に対して表現学習を利用する上での次の課題の 1 つである．

行動認識の分野で教師なしデータの活用を試みた研究として，Huynh らは，生成モデルによる教師なし学習と識別器による教師あり学習を組み合わせることを提案している [136]．具体的には，前者に Multiple-Eignvectors 法を利用して低次元の特徴量を教師なしの手法で獲得し，ラベルありのサブセットを利用した SVM による教師あり学習を行うことで認識精度が向上したことを報告している．

Huynh らの研究は生成モデルにより教師なし学習を利用するという点で興味深いものであるものの，生成モデルへの入力に人出で設計した特徴量を利用している．そのため，特徴量を人手で設計していることに関する認識精度に関する限界があると考えられる．高い表現力を持つディープネットによる教師なし学習や半教師あり学習は，特に画像認識の分野で State-of-the-art な精度を持つことが近年の研究で明らかになってきている．ディープネットに関連する代表的な教師なし，半教師あり学習手法として次のようなものが知られている．

- 制限付きボルツマンマシン（RBM）[68]
- 自己符号化器（Auto-Encoder），ノイズ除去型自己符号化器（De-Noising Auto-Encoder [137]
- 変分自己符号化器（Variational Auto-Encoder）[138] およびにその亜種
- Ladder Networks [139]
- Generative Adversarial Nets の応用 [140]

これらの手法の有効性は主に画像認識の領域で報告されているものであり，直接的にウェアラブルセンサによる行動認識においても有効であるかは自明ではないものの，これらの手法を行動認識の分野に拡張することで教師なしデータからの表現の獲得も可能になると考えられる．

ディープネットによる教師なし学習をウェアラブルセンサによる行動認識の分野に適用した代表的な研究として，Plotz らの研究がある [12]．Plotz らは，加速度の生データに対して制限付きボルツマンマシンによる教師なし学習を適用し，経験的な特徴量と認識精度による比較を行っている．ただし，既存の研究結果では，この認識精度はラベル付きデータセットが十分に存在する場合には完全な教師あり学習による CNN や LSTM の認識精度を下回ることが報告されている [19, 21]．一方，これらの検証では全データに対してラベル付きデータが付与されている設定での実験であるが，むしろ現実的には一部に教師ありデータが付いた上で，大量の教師なしデータが得られるという設定が妥当であろう．例えば，Inoue らが提供する看護センシングデータセットには，22 人の看護師行動に関する 2 週間分のラベル付きデータセット合計 346.5 時間分と，60 人の看護師に関するラベルなしデータセット 1655 人日分が含まれている．既存のデータセットの多くはラベル付きデータの収集に注

力しているものの，大量のラベル無しデータを含むデータセットにおける教師なし学習の有用性検証は認識精度改善に向けた重要な方向性である．

次に,「どのような表現を獲得するか」という What の観点から本論文の課題について考察する．認識精度という観点から見たときに本研究は，実用上必要となる学習データが少ないユーザに対する認識精度を高めることを目的にを目指した研究である．この背景には，ウェアラブルセンサによる観測値は同じ行動をしていてもユーザにより大きく異るものであり，ユーザに関するバイアスが存在するということがある．本論文ではこのユーザバイアスについてのみ取り上げたが，ウェアラブルセンサによる行動認識で問題となるバイアスとしてはその他に実験環境と実環境において認識精度が大きく低下する環境バイアスや計測に利用するデバイスの種類により生じるデバイスバイアス，設置位置バイアスなどが報告されている．これらのバイアスもユーザバイアスと同様実環境で認識システムを利用しようとする場合に避けることが難しく，また大量にデータを獲得するというアプローチでは難しいバイアスであり，これらの多様なバイアスに対して頑健な表現の獲得が行動認識では求められることとなる．

本研究では，ユーザバイアスを解消する方法として，ユーザ汎用な表現を学習する方法およびに効率的に特定のバイアスに特化した表現を少数データから効率的に学習する方法を提案した．今回の検証はユーザバイアスに対する検討のみであるものの，提案手法はバイアスがユーザによって起こることに紐づく性質を利用したものではなく，ユーザバイアスに強く特化した手法ではないため原理的には直接的にその他のバイアスに利用可能である．例えば，デバイスの種類の違いにより生じるデバイスバイアスの問題に対してはなるべく多様なデバイスによるデータから表現を獲得することでデバイス汎用な表現を学習したり，あるいは特定のデバイスに対して特化させることでさまざまなデバイスにおいて認識精度を高められるのではないかと期待される．一方，バイアスの種類により観測値への違いがどのように発生するかは異なると考えられる．例えば，ユーザの違いにより生じるバイアスはどちらかと言えば歩き方などローカルなパターンの違いであるのに対して，設置位置により生じるバイアスは重力加速度の影響の仕方といったよりグローバルなパターンの違いであると考えられる．今回ユーザバイアスについて有効性が確認されたディープネットやその適応手法をその他のバイアスの問題についても適用することにより，どのような違いに対して有効であるのかを確認していくことが必要である．

「どのような表現を獲得するか」という観点にたったときの本論文が扱っていないもう 1 つの論点に，上記のようなバイアスのない状態での認識精度を高める表現

を学習するかという論点がある．すなわち，本論文はディープネットがバイアスの少ないデータに対して良い認識精度を達成するという従来研究の知見を元に，バイアスがあるデータに対してもその認識精度をバイアスの少ないデータに対する精度に近づけようとするものである．そのため，そもそもディープネットで学習できる観測した信号から行動への写像の学習自体が十分でない場合認識精度が不十分になる可能性がある．例えば，4 章で示したようにディープネットを利用する場合でも検証データに対する 4 データセット平均の認識精度は 0.70 程度（F 値）であり，アプリケーションによっては十分ではない．観測信号や行動はダイナミックでその対応関係は確率的であり，認識精度にはそもそも限界があるが，いずれにせよディープネットにより取り出すことができる特徴の限界に伴う認識精度の限界が本論文で提案した手法には存在する．

認識精度を高めるということを考えたときに，本論文を初め多くの論文で行動を便宜上 1-hot ベクトルで表現していることがボトルネックの 1 つになっている可能性がある．すなわち，行動は実際的にはこのような離散的なカテゴリ表現で表されるものではない可能性があり，そのことが学習効率や認識精度に影響を負の影響を与えている可能性がある．例えば，歩くという行動と走ると言う行動はどちらも動的な動きである一方で座る，寝そべると言った行動は静的であり本来横一列に並ぶようなく，何らかの階層構造を持つものであると考えるのが妥当であるが，このような階層性は本研究においては考慮されていない．

行動の階層性を考慮した既存研究として，例えば，Khan らは行動を分類する際に単純に多クラス分類により分類するのではなく経験的に決定した行動分類モデルに従い例えばはじめは動いている/止まっている，動いている場合にはその姿勢が寝そべっているかどうかを分類する，寝そべっていない場合には... と順に分類を行うことで認識精度が向上することを報告している [141]．同様に，Kasteren も階層的に分類を行うことで認識精度の改善が可能であったと報告している [142]．

このような観点から見ると，そもそも最適化する損失として式 6.1 のような交差エントロピーを利用することはそもそも限界がある．すなわち，本論文を含む行動認識研究の多くはそもそも最適化しようとしている損失関数に限界があり，それゆえに生じている誤差がある可能性がある．

類似した観点として，行動の構成性という観点もある．ここで行動の構成性とは，歩くという行動は右足のかかとをつく，右足のつま先をつく，左足のかかとを蹴りだす... といったいくつかの更に小さな行動の組み合わせにより表されることを意味している．この構成性もまた，行動には階層性があることを示唆している．

これらのような行動の階層性を考慮に入れた表現を学習することは，より高精度に行動を認識したりデータが少ない環境で効率的に認識モデルを構築するうえで重要である．音素などのように人間が知覚できるものと違い行動素がどのようなものであるか自体が難しい問題であるが，獲得された表現により行動の階層性が表現されているのか，また階層性をどのように表現すれば良いのかは重要な研究課題である．

7.2.2 プライバシーへの配慮に対する本論文の貢献と課題

本研究の結果はウェアラブルセンサにより観測される生体データのようなパーソナルデータに対してディープネットを応用する研究として位置づけられる．4 章で示された特徴量のユーザ依存性は，通常利用されているディープネットを単純に適用すると，本来収集したかった行動以外の情報を暗黙的に収集してしまう可能性があることを示唆するものである．4 章で明確に示されたユーザの依存性それ自体が直接プライバシーを侵害するかどうかは自明ではないものの，この結果は少なくとも意図しない情報を収集してしまうリスクを示す結果である．パーソナルデータに対して表現学習をプライバシーに配慮した形で利用するためには，認識精度を高めるという観点のみを考慮した表現ではなく意図しない情報を取得するリスクを低減するような表現を学習する必要がある．

平成 15 年に成立して以来 12 年ぶりに改定された個人情報保護新法によれば，その新 17 条 2 項において，要配慮個人情報を本人の同意なく取得することを禁じている．ここで，要配慮個人情報とは新 2 条 3 項によれば,「本人の人種，心情，社会的身分，病歴，犯罪の経歴，犯罪により被害を被った事実その他本人に対して不当な差別，偏見，その他の不利益が生じないようにその取扱に特に配慮を要するものとして政令で定める記述等が含まれる個人情報」として定義されている．例えば病歴は歩き方や行動パターンと密接に関連していると考えられ，ウェアラブルセンサにより観測された生体データは暗黙的にこのような配慮が必要な情報を含んでいる可能性がある．本人の同意なく取得することを禁じるという文言をどう解釈するかによるが，要配慮情報を収集しないようにするという立場からすると，意図しない情報を収集してしまうリスクは可能な限り低減することが望ましい．本研究は，学習される表現が特定の情報を持たなくするようにする手法を提案することにより，プライバシーへの配慮が必要なパーソナルデータに対する表現学習の適用可能性を高めることに貢献するものである．

このような情報が予測できてしまうことに関するプライバシーリスクに関する指

摘は，2 章に述べたように，ディープネットのようなアルゴリズムが存在する以前から行われてきた．この問題に対する従来の解決策は大別して，1. 情報が取得されうるリスクに応じてセンサを利用しないようにする，2. 取得するべきでない情報を予測できないように生データを加工する，という手法に分けることができる．5 章で行ったユーザを識別できるような特徴量を利用しないようにするというアプローチは，大別すると 2 の枠組みの 1 手法として捉えることができる．本研究と同様に 2 のアプローチを取ることを提案しているその他の研究として，Chakraborty らが提案している特徴量の難読化というアプローチがある [67]．具体的には，特徴量をすべて使うのではなく本当に予測したい (あるいは予測することをユーザが同意した) 情報の予測は可能だが機微情報の予測は難しくなるような特徴量のみを利用することでユーザが望まない情報を推定されるリスクを抑えることを提案している．本論文の主たる目的はディープネットがウェアラブルセンサによる行動認識の精度向上に寄与することを確認した上で，その実用上の課題を解決する手法を提案することであるため本論文中では Chakraborty らの研究と明確な比較実験を行っていないものの，本論文で提案した手法は Chakraborty らの研究と比べていくつかの明確な利点がある．

具体的には，まず Chakraborty らの研究では上記を特徴選択の枠組みで実現することを提案しているため，認識精度の観点で表現学習を利用している提案手法が優れていると考えられる．すなわち，Chakraborty らの提案は従来的な特徴設計で得られた特徴量の集合の中から行動の識別性と取得したくない情報の不識別性の双方を考慮して特徴量を選択する方法であり，そもそもの特徴量の集合を人間が設計していることに紐づく認識精度の限界がある．実際，4 章で示したように特徴設計により得られた特徴は提案手法が利用している表現学習によるものと比較して認識精度が低く，この点で提案手法に優位性があることを示唆している．また，Chkaraborty らの研究は特徴選択を行うという枠組みの提案であり，具体的にどのように特徴選択を行えば良いのかという部分について提案されておらず，特徴量を変更した場合に認識したい対象の認識精度と認識したくない対象の認識精度がどのように変化するかを検証するにとどまっている．このように，5 章の提案手法は Chakraborty らが提案した特徴量の難読化というアプローチを表現学習の文脈で実現する手法であり，本論文の結果は，生データからさまざまな情報が推測されうるというウェアラブルセンサを利用した行動認識における問題に対して表現学習という観点からアプローチすることの有効性を示したものである．

プライバシーに関する技術的対策を進めていく上では，逆説的であるがプライバ

シーを侵害するような技術の開発が必要となる．すなわち，プライバシーを侵害するような強い攻撃者を想定した上で，その攻撃者から情報を適切に守るような手法を開発していく必要がある．また，そのためにも5章で述べたような大規模かつユーザの属性情報を含むようなデータセットを構築することが必要である．大規模なデータセットによる実験を通じて，ウェアラブルセンサによる生体データからどのような情報が取得可能であるかの知見を蓄積していく必要がある．

一方で，プライバシーの保護技術は基本的にある攻撃者を想定した上で議論されるものであり，想定を超える攻撃者に対しては無力である．しかし，アルゴリズムによる認識精度が発展し続けていることに代表されるように，攻撃者の能力は日々高まっており，このような無限の可能性を持つ攻撃者を想定した技術的対策を及ぼすことは残念ながら難しいか，ほぼ不可能であろう．例えば，5章では学習時に利用する仮想的な敵対的分類器の分類能力が実際の敵対的分類器より低い場合には，依然として特徴量からユーザを識別できてしまう可能性があることを示していた．また，より一般的な例ではプライバシーというよりむしろセキュリティと関連が深い技術であるが，RSA暗号は素因数分解が現在のコンピューティング能力では困難であることに立脚しているが，例えば量子コンピュータが実現すればその推定を行うことは容易であることは広く知られている．そのため，プライバシーへの配慮を適切に行うためには技術的な対策だけでなく法的な対策を適切に行なっていく必要がある．

プライバシーへの配慮に関する法的な動向として，本項の冒頭部分でも少し触れたが，平成28年1月1日より個人情報保護法の改定版が一部施工された．改定の背景にはいくつかの理由があるが，そのうちの1つとして従来法の個人情報保護法にいて個人情報として取り扱う範囲の曖昧さ（グレーゾーン）が技術進展により拡大していることがある．このような背景から，本人の同意がなくともデータを利活用可能とする枠組みが導入されたり，個人情報の範囲が明確化されたりと，個人の権利利益に損害が生じない範囲でグレーゾーンを減らすことによりパーソナルデータの利活用を推し進めることを狙った改正となっている．

この法整備の観点から見たときに本論文が示したような中間表現に意図しない情報が現れる可能性は，パーソナルデータに対するディープネットの適用におけるグレーゾーンを生じさせる可能性がある．例えば，新17条2項において，要配慮個人情報を本人の同意なく取得することを禁じている（具体的には，条例では「個人情報取扱事業者は、次に掲げる場合を除くほか、あらかじめ本人の同意を得ないで、要配慮個人情報を取得してはならない」と記述されている）．このときに，ここま

で議論したように，最終的に行動を識別することを目的にするようなネットワークであったとしてもその中間表現では意図しない情報 (すなわち要配慮情報) を収集している可能性があることをどのように捉えるかは曖昧な問題と思われる．あるいは，学習したモデル自体を個人情報としてみなして保護の対象とすべきかも曖昧であろう．これらの議論は特にディープネットを適用する場合に実務的に問題となる領域であると考えられ，意図しない情報を取得しないようにする技術開発を含めて，技術と規制双方の観点から検討を進めていく必要がある．

7.3 ウェアラブルセンサによる行動認識技術の活用可能性とその課題

本論文はプライバシーおよびに認識精度の双方の観点から優れた表現を学習するディープネットの拡張手法の提案により，ウェアラブルセンサによる行動認識において高い認識精度を達成するディープネットを行動認識に対して適用できるようにしたものである．ウェアラブルセンサによる行動認識により得られるユーザの実世界における行動データは従来は大規模に取得が困難であった情報であり，これらの情報を活用することにより 1 章でも述べたヘルスケアシステムのような新たなサービスや事業が創発される可能性がある．本節ではウェアラブルセンサによる行動認識技術をアプリケーションで活用するという観点から，まずウェアラブルセンサを取り巻く環境について整理することでその活用可能性について述べたのちに，活用に向けた課題について考察する．

7.3.1 ウェアラブルセンサを取り巻く環境と将来の展望

ウェアラブルデバイスの普及・多様化

2 章にも記述したようにユーザが持ちあるくセンサによりそのユーザの行動を認識する技術については 1990 年代後半から研究されていたが，従来このようなセンシング技術はセンサのハードウェアレベルでの実装を必要としたため比較的小規模なものに限られてきた．例えば，2005 年頃に MIT が開発した LiveNet とよばれるシステムは，専用デバイスを構築することでセンシングを行っていた．それに対して現在では多くの人がスマートフォンを持ち歩いていることやように，多くの人がセンサを持ち歩いている．このようなすでにユーザが所有しているセンサを利用するこ

とで，従来と比較して劇的に低コストに人間行動をセンシングできる環境が整いつつある．

また，今後はスマートフォンだけでなくユーザが持ち歩くさまざまなものにセンサが導入されていくものと考えられる．例えば，すでに市場に現れている例で言えばリストバンド型のセンサやスマートウォッチがある．また，Apple は 2014 年にはイヤホンによる行動モニタの米国特許を取得しており [1]，2016 年 9 月に加速度センサを内蔵したイヤホンを AirPods を発表している．このようなデバイスはヒアラブルデバイス (hearable device) と呼ばれ，ウェアラブルデバイスの発展技術として期待されている．その他にも，衣服や靴などユーザが身につけるさまざまな物にセンシング機能が搭載されるようになっている．

デバイスや設置位置の多様化に伴い計測可能な情報の種類が増加すると考えられる．例えば，スマートフォンでは加速度センサやジャイロスコープが行動認識を行う上では主流であるが，スマートウォッチやリストバンド型のセンサを利用する場合は心拍センサなどが加えて利用可能である．このように，今後は多様なデバイスの普及に伴いより精緻なユーザの情報が集められると想定される．

第 5 世代移動通信システム

デバイスの普及に加えて，通信網の発展もウェアラブルセンサにより行動をセンシングする上では必要である．通信網に関しては，2020 年ごろの実用化を目指して第 5 世代移動通信システム，いわゆる 5G の研究開発が進められている．5G とは，現行の第 4 世代通信システム（4G あるいは LTE）に代わる新たな通信システムであり，現行の LTE の 100 倍となる 10Gbps の最高伝送速度とすることや，膨大な数のセンサ端末が接続することを想定して 100 万台/km^2（現行 LTE の 100 倍）の接続機器数に対応することが主要性能の 1 つとして盛り込まれている [2]．このような通信システムが実用化されることにより，センシングした大量のデータを通信し処理するようなサービスを実現することが可能となり，新規市場の創出が期待されている．

現在はいまだセンサの普及およびにデータの蓄積段階であるが，デバイスや通信網が整備されれば，ウェアラブルセンサにより人間行動を計測するハードウェア的な制約は低減する．結果として，将来的には実世界における人間行動に関するデー

[1] USP: 8655004，“Sports monitoring system for headphones, earbuds and/or headsets”

[2] 具体的な数値については総務省発行の電波政策 2020 懇談会報告書 `http://www.soumu.go.jp/main_content/000423012.pdf` を参考にした．

タがさまざまなセンサにより計測され蓄積されるようになる，あるいはそのハードルが大きく低下すると想定される．

7.3.2 ウェアラブルセンサによる行動認識の活用に向けた課題

センシングデバイスや通信環境双方の観点で人間行動を計測するハードルが低下するに伴い，収集されたデータから情報を引き出すためには分析技術の重要性は高まると考えられる．ウェアラブルセンサによるデータから情報を引き出すための分析技術の重要な観点の１つが本論文で扱ったような認識精度の改善であり，またプライバシーへの配慮である．一方，ウェアラブルセンサによる行動認識を実際にアプリケーションとして活用する場合に，単にデータから情報を引き出すだけでなく，情報をユーザに価値のある形で還元し，ユーザの継続利用を通してユーザの新たなデータ蓄積しより良い（新しい，あるいは高精度な）情報/価値に変換するというサイクルを構築することが必要となる．

例えば，以降日々どのような行動をしたかを認識することにより健康に関するアドバイスを行う健康アドバイザーシステムを考えてみる．また，このシステムは技術要素としてユーザが設置したリストバンド型センサによるデータから行動認識を利用しており，ユーザの行動履歴情報をシステムの情報として利用しているとする．

情報をユーザに価値のある形で還元するという観点から考えると，認識された行動の系列を単に表示することもユーザに対して気づきを与える可能性があるものの，例えば自分と類似したユーザと比較して自分はどのくらい歩いているのかと言った情報はユーザにとってより意味のある情報になるだろう．あるいは，一歩進んで例えば「焼肉を食べたので今日はエスカレーターを使わず階段を使うと良いでしょう」といった行動を推薦することができればシステムを利用する価値が高まると考えられる．このように，ウェアラブルセンサから推定された情報を価値に還元するためには，本論文が行った高精度な認識に加えて何らかの方法でユーザに価値のある形に変換するというプロセスが必要となる．

上記のように引き出された情報を適切な形で直接フィードバックすることで価値として還元するような方法だけでなく，金銭的なメリットを還元するというような方法も考えられる．例えば，実世界で金銭的なメリットをユーザに還元している例として健康増進型保険がある．健康増進型保険とは，ウェアラブルセンサなどによる種々の情報から計測されるユーザの行動による健康度合いに応じて保険料の値引きなどさまざまな付加価値が得られるタイプの保険である．最も代表的な商品であ

るディスカバリー社が提供する「バイタリティー・プログラム」では，保険加入者のジム利用や1日あたりの歩数などに加えて，ウェアラブル機器を利用した日常的な健康増進活動の実行状況の記録に利用している（なお，ディスカバリー社は2016年7月22日に住友生命とソフトバンクと連盟で健康増進保険の開発プロジェクトの立ち上げを発表している[3]）．保険加入者はデータを提供する代わりに健康増進行動のとり方に応じたポイントやステータスに応じて保険料割引や提携パートナー企業から特典を得られるなどのメリットを得られるという仕組みである．逆に保険事業者は健康に関するデータが得られることに加え，契約者が健康になることで払う保険料が低減するため，双方にとってメリットがある．

このような情報を価値に変換するプロセスは，ユーザに対して価値を届けるという意味だけでなくサステナブルにデータを集めることにつながるという点でも重要である．例えば，健康アドバイザーを多くのユーザが利用すればより信頼度の高い統計情報や，類似ユーザに関する情報などを提供することができるようになる．あるいは，あるユーザに長期間利用してもらうことができれば行動変容の情報や健康の変化に関する情報を得ることができる可能性がある．このように，ユーザに対して価値を届けることで継続利用を促し，またより価値のある情報を創出するというプロセスがウェアラブルセンサによる行動認識を活用する上では重要である．

先程あげた健康増進型保険の異なる例として，2016年2月8日には損保ジャパン日本興亜ひまわり生命は「健康増進に向けた共同ビジネスの展開の開始について」というニュースリリースの中でウェアラブル端末の提供を行うフィットビット社と共同でウェアラブル端末を利用した健康増進プロジェクトを発表している[4]．ニュースリリースの中に「健康保険組合との連携のもと、社員の健康情報とフィットビット社製ウェアラブル端末を通じて収集できる活動データをもとに、疾病と活動データの因果関係を分析する取組み。」検討していくとあるように，一旦データが蓄積されればそのデータを分析することにより従来はわからなかったような知見を獲得することが可能になり，また新たなサービスへと繋がる好循環が生まれていくであろう．

本論文はウェアラブルセンサによる行動認識のための表現学習技術という，ウェアラブルセンサによるデータを適切な方法でより良い情報に変換する技術を開発するものであり，実際にユーザに対して価値を還元する体系的な方法を開発するもの

[3]プレスリリース URL: http://www.softbank.jp/corp/group/sbm/news/press/2016/20160721_03/

[4]ニュースリリース URL: http://www.himawari-life.co.jp/unique/images/company/news/a-01-2016-02-08.pdf

ではない点を言及しておく．例えば，本論文で扱った認識精度の改善は行動認識を行ううえで必須の観点であるものの，どの程度の認識精度が達成されるかは構築したいシステムによるものであり，表現学習により得られる情報が価値につながりうるかや，つなげるためにどのような仕組みが必要となるかは個別のアプリケーション領域での検討が必要となる．

オンライン世界において大量のデータを集めている Google が AI の技術開発において先行しているように，データの価値は特にデータから価値を創発する技術の進歩に伴って重要となっている．特に，本論文が扱ったディープネットの有効性はデータが大きいほど高まるものであり，大規模データの価値は従来よりも相対的に高まっている．逆に言えば，データの獲得競争で遅れを取った場合には，同時に技術開発でも遅れを取ることになる可能性がある．実務的には，データの収集と技術開発を並行で進めていくことで，実世界における人間行動と紐付いた知見の獲得やヘルスケア，コンテキストアウェアなどの高度な認識を伴うシステムを実用化に結びつけていくサイクルを構築することが必要である．

第8章　結論

1章で，研究背景およびに研究目的を整理し，2章でウェアラブルセンサを利用した行動認識の既存研究を整理し本研究が取り組む問題の意義およびに既存の研究について説明した．3章でウェアラブルセンサを利用した行動認識に対して表現学習を適用した既存研究を整理し，4章から6章の位置づけを整理した．

4章から6章では，表現学習の利用およびに本論文で開発した提案手法により，実世界の人間行動をより正確に，プライバシーに配慮した形で計測できることを確認した．まず，4章で110万サンプル以上のウェアラブルセンサによる行動認識データに対して最大24層のディープネット適用し，経験的に利用されてきた特徴量より優れた認識精度を達成することをユーザ独立認識精度において確認した．これにより，学習データに複数人のユーザのデータを用意した場合にディープネットが学習する特徴量は単に学習ユーザに過学習したようなものではなく，ユーザ汎化性能を持つものであることが確認された．一方，学習された特徴空間上でのクラスタリング結果はユーザが過度に特定のクラスタに固まる傾向を示しており，ユーザに関する情報は完全に取り除かれるわけではないことを確認した．この結果は，ディープネットを適用した場合に中間表現が意図しない情報を持つリスクが有ることを示唆している．

5章では，4章で生じた問題を解決するために，学習される特徴量がユーザに関して独立になるような，すなわちユーザに関する情報を含まないような特徴を学習するようにディープネットを拡張する方法を提案した．実データを利用した実験により既存のディープネットによる手法と提案手法を比較し，提案手法により認識精度を維持しながらユーザの情報を特徴量から取り除くことができることを確認した．これにより，行動の予測性能をの低下を最小限にとどめながら，ユーザやユーザの属性情報を収集してしまうリスクを抑制することができる．

6章では，特定ユーザに対する少数のラベル付きデータと多数のラベルなしデータを利用して，複数のユーザのデータにより事前に学習された非個人認識モデルを特定のユーザに対して効率的に適応する方法を提案した．複数のベースライン手法と提案手法の認識精度改善度合いを比較し，既存の手法では改善困難なラベル付き

データ数が極めて少数な場合でも提案手法により認識精度が改善可能であることなどを確認した．これにより，ディープネットによる表現学習がウェアラブルセンサによる行動認識に対して有用であることを示した．

以上により，本論文ではウェアラブルセンサによる行動認識に対して有効な表現学習技術の開発を目的に，本タスクにおいて論点となる認識精度とプライバシーへの配慮の観点を考慮した表現学習手法を提案し，検証を通じて有効性を確認した．また結果として，ディープネットによる表現により認識精度が従来法と比べて高まること，望まない情報を取得する可能性を低減するという意味でプライバシーに配慮するように表現学習を拡張できること，また学習データが少ないユーザの行動を認識する必要がある場合でも他ユーザで学習された表現を修正するユーザ適応技術によって従来より高い認識精度が達成できることを確認し，ディープネットによる表現学習がウェアラブルセンサによる行動認識に適用可能であることを示した．

一方で，行動計測を現実的に行うための課題は，7 章に示したように依然として存在している．これらの問題を機械学習，情報理論の立場からアプローチすることで，人間行動の計測の技術基盤の構築に貢献していきたいと考えている．

参考文献

[1] B. H. Dobkin and A. Dorsch, "The Promise of mHealth: Daily Activity Monitoring and Outcome Assessments by Wearable Sensors," *Neurorehabilitation and Neural Repair*, vol. 25, pp. 788–798, nov 2011.

[2] M. Sung, C. Marci, A. Pentland, C. Tavormina, G. Abowd, N. Holter, J. Generelli, R. Bashshur, J. Sanders, G. Shannon, V. Kirk, S. Bohn, W. Flemons, J. Remmers, R. DeVaul, M. Sung, J. Gips, A. Pentland, C. Elledge, E. Tapia, N. Marmasse, S. Intille, K. Larson, M. Sung, S. Basu, M. Sung, J. Weaver, D. Klapper, L. Bao, K. V. Laerhoven, A. Schmidt, H. Gellersen, N. Eagle, A. Pentland, S. Intille, C. Kukla, X. Ma, R. Kessler, P. Berglund, O. Demler, R. Jin, D. Koretz, K. Merikangas, A. Rush, E. Walters, P. Wang, N. Marmasse, J. Fogarty, and V. Gerasimov, "Wearable Feedback Systems for Rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 2, no. 1, p. 17, 2005.

[3] B. G. Steele, B. Belza, K. Cain, C. Warms, J. Coppersmith, and J. Howard, "Bodies in Motion: Monitoring Daily Activity and Exercise With Motion Sensors in People With Chronic Pulmonary Disease.," *Journal of rehabilitation research and development*, vol. 40, no. 5 Suppl 2, pp. 45–58, 2003.

[4] M. G. Benedetti, A. Di Gioia, L. Conti, L. Berti, L. Degli Esposti, G. Tarrini, N. Melchionda, and S. Giannini, "Physical Activity Monitoring in Obese People in the Real Life Environment.," *Journal of neuroengineering and rehabilitation*, vol. 6, p. 47, 2009.

[5] O. Amft and G. Tröster, "Recognition of Dietary Activity Events Using On-Body Sensors.," *Journal of Artificial intelligence in medicine*, vol. 42, pp. 121–36, feb 2008.

[6] E. Ertin, N. Stohs, S. Kumar, A. Raij, M. Al'Absi, and S. Shah, "AutoSense: Unobtrusively Wearable Sensor Suite for Inferring the Onset, Causality, and Consequences of Stress in the Field," in *ACM Conference on Embedded Networked Sensor Systems (SenSys)*, (New York, New York, USA), pp. 274—-287, ACM Press, 2011.

[7] R. J. Robles and T.-H. Kim, "Review: Context Aware Tools for Smart Home Development," *International Journal of Smart Home*, vol. 4, no. 1, 2010.

[8] S. Inoue, N. Ueda, Y. Nohara, and N. Nakashima, "Mobile Activity Recognition for a Whole Day: Recognizing Real Nursing Activities With Big Dataset," in *ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, (Osaka, Japan), pp. 1269–1280, ACM Press, 2015.

[9] Y. Iwasawa and I. E. Yairi, "Life-Logging of Wheelchair Driving on Web Maps for Visualizing Potential Accidents and Incidents," in *Pacific Rim International Conference on Articifial Intelligence (PRICAI)*, pp. 157–169, Springer, 2012.

[10] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 8, pp. 1798–1828, 2013.

[11] N. Y. Hammerla, R. Kirkham, P. Andras, and T. Ploetz, "On Preserving Statistical Characteristics of Accelerometry Data Using Their Empirical Cumulative Distribution," in *International Symposium on Wearable Computers*, pp. 65–68, ACM, 2013.

[12] T. Plötz, N. Y. Hammerla, and P. Olivier, "Feature Learning for Activity Recognition in Ubiquitous Computing," in *International joint conference on Artificial intelligence (IJCAI)*, p. 1729, 2011.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet Classification With Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 1097–1105, 2012.

[14] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the Gap to Human-Level Performance in Face Verification," in *Proceedings of the*

IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708, 2014.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *arXiv preprint arXiv:1512.03385*, 2015.

[16] G. Hinton, L. Deng, D. Yu, G. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," *IEEE Signal Processing Magazine*, vol. 29, pp. 82–97, nov 2012.

[17] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-Level Control Through Deep Reinforcement Learning," *Nature*, vol. 518, pp. 529–533, feb 2015.

[18] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the Game of Go With Deep Neural Networks and Tree Search," *Nature*, vol. 529, pp. 484–489, jan 2016.

[19] J. B. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy, "Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition," in *International joint conference on Artificial intelligence (IJCAI)*, 2015.

[20] F. J. Ordóñez and D. Roggen, "Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.

[21] N. Y. Hammerla, S. Halloran, and T. Ploetz, "Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables," *International joint conference on Artificial intelligence (IJCAI)*, 2016.

[22] F. Foerster, M. Smeja, and J. Fahrenberg, "Detection of Posture and Motion by Accelerometry: A Validation Study in Ambulatory Monitoring," *Journal of Computers in Human Behavior*, vol. 15, no. 5, pp. 571–583, 1999.

[23] O. Durmaz Incel, M. Kose, and C. Ersoy, "A Review and Taxonomy of Activity Recognition on Mobile Phones," *BioNanoScience*, vol. 3, no. 2, pp. 145—-171, 2013.

[24] O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition Using Wearable Sensors," *Communications Surveys & Tutorials, IEEE*, vol. 15, no. 3, pp. 1192–1209, 2013.

[25] A. Bulling, U. Blanke, and B. Schiele, "A Tutorial on Human Activity Recognition Using Body-Worn Inertial Sensors," *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, p. 33, 2014.

[26] G. Cho, S. Lee, and J. Cho, "Review and Reappraisal of Smart Clothing," *International journal of human-computer interaction*, pp. 582—-617, 2009.

[27] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine," in *International Workshop on Ambient Assisted Living*, pp. 216–223, Springer Berlin Heidelberg, 2012.

[28] J. Guo, X. Zhou, Y. Sun, G. Ping, G. Zhao, and Z. Li, "Smartphone-Based Patients' Activity Recognition by Using a Self-Learning Scheme for Medical Monitoring," *Journal of medical systems*, vol. 40, no. 6, pp. 1–14, 2016.

[29] J. K. Aggarwal and M. S. Ryoo, "Human Activity Analysis: A Review," *ACM Computing Surveys (CSUR)*, vol. 43, no. 3, p. 16, 2011.

[30] J. Pansiot, D. Stoyanov, D. McIlwraith, B. P. Lo, and G. Z. Yang, "Ambient and Wearable Sensor Fusion for Activity Recognition in Healthcare Monitoring Systems," in *4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007)*, pp. 208–212, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[31] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, "A Review of Wearable Sensors and Systems With Application in Rehabilitation.," *Journal of neuroengineering and rehabilitation*, vol. 9, p. 21, 2012.

[32] S.-W. Seon-Woo Lee and K. Mase, "Activity and Location Recognition Using Wearable Sensors," *IEEE Pervasive Computing*, vol. 1, pp. 24–32, jul 2002.

[33] T. Gu, Z. Wu, X. Tao, H. K. Pung, and J. Lu, "Epsicar: An Emerging Patterns Based Approach to Sequential, Interleaved and Concurrent Activity Recognition," in *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 1–9, IEEE, 2009.

[34] L. Bao and S. S. Intille, "Activity Recognition From User-Annotated Acceleration Data," in *International Conference on Pervasive Computing*, pp. 1–17, Springer Berlin Heidelberg, 2004.

[35] A. K. Bourke, J. V. O'brien, and G. M. Lyons, "Evaluation of a Threshold-Based Tri-Axial Accelerometer Fall Detection Algorithm," *Gait & posture*, vol. 26, no. 2, pp. 194–199, 2007.

[36] D. W. Kang, J. S. Choi, J. W. Lee, S. C. Chung, S. J. Park, and G. R. Tack, "Real-Time Elderly Activity Monitoring System Based on a Tri-Axial Accelerometer," *Disability and Rehabilitation: Assistive Technolog*, vol. 5, no. 4, pp. 247–253, 2010.

[37] N. D. Lane, Y. Xu, H. Lu, S. Hu, T. Choudhury, A. T. Campbell, and F. Zhao, "Enabling Large-Scale Human Activity Inference on Smartphones Using Community Similarity Networks (Csn)," in *International conference on Ubiquitous computing (UbiComp)*, (New York, New York, USA), p. 355, ACM Press, 2011.

[38] Q. Li, J. Stankovic, M. Hanson, A. T. Barth, J. Lach, G. Zhou, and Others, "Accurate, Fast Fall Detection Using Gyroscopes and Accelerometer-Derived Posture Information," in *International Workshop on Wearable and Implantable Body Sensor Networks*, pp. 138–143, IEEE, 2009.

[39] S. Morris and J. Paradiso, "Shoe-Integrated Sensor System for Wireless Gait Analysis and Real-Time Feedback," in *Annual Conference and the Annual Fall Meeting of the Biomedical*, vol. 3, pp. 2468–2469, IEEE, 2002.

[40] L. Atallah, B. Lo, R. King, and G.-Z. Yang, "Sensor Positioning for Activity Recognition Using Wearable Accelerometers," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 5, pp. 320–329, aug 2011.

[41] M. Sung, R. DeVaul, S. Jimenez, J. Gips, and A. Pentland, "Shiver Motion and Core Body Temperature Classification for Wearable Soldier Health Monitoring Systems," in *International Symposium on Wearable Computers (ISWC)*, vol. 1, pp. 192–193, IEEE, 2004.

[42] A. A. Ali, S. M. Hossain, K. Hovsepian, M. M. Rahman, K. Plarre, and S. Kumar, "mPuff: Automated Detection of Cigarette Smoking Puffs From Respiration Measurements," in *International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 269–280, IEEE, apr 2012.

[43] P. H. Veltink, H. B. Bussmann, W. de Vries, W. L. Martens, and R. C. Van Lummel, "Detection of Static and Dynamic Activities Using Uniaxial Accelerometers.," *IEEE transactions on rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 4, pp. 375–85, dec 1996.

[44] M. Sekine, T. Tamura, M. Ogawa, T. Togawa, and Y. Fukui, "Classification of Acceleration Waveform in a Continuous Walking Record," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 3, pp. 1523–1526, IEEE, 1998.

[45] M. Sekine, T. Tamura, M. Akay, T. Fujimoto, T. Togawa, and Y. Fukui, "Discrimination of Walking Patterns Using Wavelet-Based Fractal Analysis.," *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 10, pp. 188–96, sep 2002.

[46] Y. Liu, R. Jafari, C. Lu, S.-I. Institute for Computer Sciences, SIGCHI (Group : U.S.), and ACM Digital Library., "A Feature Selection-Based Framework for

Human Activity Recognition Using Wearable Multimodal Sensors," in *International Conference on Body Area Networks*, pp. 92—-98, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2011.

[47] B. Nham, K. Siangliulue, and S. Yeung, "Predicting Mode of Transport From iPhone Accelerometer Data," *Machine Learning Final Projects, Stanford University*, 2008.

[48] G. M. Weiss and J. W. Lockhart, "The Impact of Personalization on Smartphone-Based Activity Recognition," in *AAAI Workshop on Activity Context Representation: Techniques and Languages*, 2012.

[49] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford, "A Hybrid Discriminative/Generative Approach for Modeling Human Activities," 2005.

[50] N. Ogawa, K. Kaji, and N. Kawaguchi, "Effects of Number of Subjects on Activity Recognition -HASC2010corpus," in *International Workshop on Frontiers in Activity Recognition using Pervasive Sensing (IWFAR)*, pp. 48—-51, 2011.

[51] K. Forster, S. Monteleone, A. Calatroni, D. Roggen, and G. Troster, "Incremental kNN Classifier Exploiting Correct-Error Teacher for Activity Recognition," in *International Conference on Machine Learning and Applications*, pp. 445–450, IEEE, dec 2010.

[52] J. Parkka, L. Cluitmans, and M. Ermes, "Personalization Algorithm for Real-Time Activity Recognition Using PDA, Wireless Motion Bands, and Binary Decision Tree," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, pp. 1211–1215, sep 2010.

[53] Z. Zhao, Y. Chen, J. Liu, Z. Shen, and M. Liu, "Cross-People Mobile-Phone Based Activity Recognition," in *International joint conference on Artificial intelligence (IJCAI)*, vol. 11, pp. 2250–2545, Citeseer, 2011.

[54] H. Hachiya, M. Sugiyama, and N. Ueda, "Importance-Weighted Least-Squares Probabilistic Classifier for Covariate Shift Adaptation With Application to Human Activity Recognition," *Neurocomputing*, vol. 80, pp. 93–101, 2012.

[55] A. Reiss and D. Stricker, "Personalized Mobile Physical Activity Recognition," in *International symposium on wearable computers (ISWC)*, pp. 25–28, ACM, 2013.

[56] A. Raij, A. Ghosh, S. Kumar, and M. Srivastava, "Privacy Risks Emerging From the Adoption of Innocuous Wearable Sensors in the Mobile Environment," in *Annual conference on Human factors in computing systems - CHI '11*, (New York, New York, USA), p. 11, ACM Press, 2011.

[57] M. Haris, H. Haddadi, and P. Hui, "Privacy Leakage in Mobile Computing: Tools, Methods, and Characteristics," *arXiv preprint arXiv:1410.4978*, 2014.

[58] A. Stopczynski, R. Pietri, A. Pentland, D. Lazer, and S. Lehmann, "Privacy in Sensor-Driven Human Data Collection: A Guide for Practitioners," *arXiv preprint arXiv:1403.5299*, 2014.

[59] J. Jun Han, E. Owusu, L. T. Nguyen, A. Perrig, and J. Zhang, "ACComplice: Location Inference Using Accelerometers on Smartphones," in *International Conference on Communication Systems and Networks (COMSNETS 2012)*, pp. 1–9, IEEE, jan 2012.

[60] S. Chakraborty, C. Shen, K. R. Raghavan, Y. Shoukry, M. Millar, and M. Srivastava, "ipShield: A Framework for Enforcing Context-Aware Privacy," in *Workshop on Secure Data Management*, pp. 85—-100, [Association for Computing Machinery], 2014.

[61] S. Brdar, D. Ćulibrk, and V. Crnojević, "Demographic Attributes Prediction on the Real-World Mobile Data," in *Proc. Mobile Data Challenge by Nokia Workshop, in Conjunction with International Conference on Pervasive Computing,*, (Newcastle, UK), 2013.

[62] H. Itay and Asaf Shabatai, *Noise Reduction of Mobile Sensors Data in the Prediction of Demographic Attributes.* IEEE Press, 2015.

[63] E. Owusu, J. Han, S. Das, A. Perrig, and J. Zhang, "ACCessory: Password Inference Using Accelerometers on Smartphones," in *Workshop on Mobile Computing Systems & Applications - HotMobile '12*, (New York, New York, USA), p. 1, ACM Press, 2012.

[64] H. Choi, S. Chakraborty, Z. M. Charbiwala, and M. B. Srivastava, "SensorSafe: A Framework for Privacy-Preserving Management of Personal Sensory Information," in *Workshop on Secure Data Management*, pp. 85–100, Springer Berlin Heidelberg, 2011.

[65] A. Cavoukian, "Privacy by Design The 7 Foundational Principals."

[66] K. R. Raghavan, S. Chakraborty, M. Srivastava, and H. Teague, "OVERRIDE: A Mobile Privacy Framework for Context-Driven Perturbation and Synthesis of Sensor Data Streams," in *International Workshop on Sensing Applications on Mobile Phones - PhoneSense*, (New York, New York, USA), pp. 1–5, ACM Press, 2012.

[67] C. Supriyo, R. Kasturi Rangan, S. Mani B., and B. Catschik, "Balancing Value and Risk in Information Sharing Through Obfuscation," in *International Conference on Information Fusion (FUSION)*, p. 2624, IEEE, 2012.

[68] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A Fast Learning Algorithm for Deep Belief Nets," *Neural Computation*, vol. 18, pp. 1527–1554, jul 2006.

[69] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 2672–2680, 2014.

[70] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning With Deep Convolutional Generative Adversarial Networks," *arXiv preprint arXiv:1511.06434*, 2015.

[71] E. L. Denton, S. Chintala, R. Fergus, and Others, "Deep Generative Image Models Using a Laplacian Pyramid of Adversarial Networks," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 1486–1494, 2015.

[72] A.-r. Mohamed, D. Yu, and L. Deng, "Investigation of Full-Sequence Training of Deep Belief Networks for Speech Recognition," in *Annual Conference of the International Speech Communication Association*, pp. 2846—-2849, 2010.

[73] A.-r. Mohamed, G. E. Dahl, and G. Hinton, "Acoustic Modeling Using Deep Belief Networks," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, pp. 14–22, jan 2012.

[74] A.-R. Mohamed, G. Hinton, and G. Penn, "Uunderstanding How Deep Belief Networks Perform Acoustinc Modeling," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

[75] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech Recognition With Deep Recurrent Neural Networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6645–6649, IEEE, may 2013.

[76] A. Graves and N. Jaitly, "Towards End-to-End Speech Recognition with Recurrent Neural Networks," in *International Conference on Machine Learning (ICML)*.

[77] Z.-H. Ling, L. Deng, and D. Yu, "Modeling Spectral Envelopes Using Restricted Boltzmann Machines and Deep Belief Networks for Statistical Parametric Speech Synthesis," in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, pp. 2129–2139, oct 2013.

[78] S. Kang, X. Qian, and H. Meng, "Multi-Distribution Deep Belief Network for Speech Synthesis," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8012–8016, IEEE, may 2013.

[79] H. Ze, A. Senior, and M. Schuster, "Statistical Parametric Speech Synthesis Using Deep Neural Networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7962–7966, IEEE, may 2013.

[80] "Wavenet: A Generative Model for Raw Audio," 2016.

[81] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 22, no. 10, pp. 1345–1359, 2010.

[82] X. Glorot, A. Bordes, and Y. Bengio, "Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach," in *International Conference on Machine Learning (ICML)*, pp. 513–520, 2011.

[83] M. Chen, Z. Xu, K. Weinberger, and F. Sha, "Marginalized Denoising Autoencoders for Domain Adaptation," in *International Conference on Machine Learning (ICML)*, 2012.

[84] C. Vollmer, H.-M. Gross, and J. P. Eggert, "Learning Features for Activity Recognition With Shift-Invariant Sparse Coding," in *International Conference on Artificial Neural Networks*, pp. 367–374, Springer, 2013.

[85] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang, "Convolutional Neural Networks for Human Activity Recognition Using Mobile Sensors," in *International Conference on Mobile Computing, Applications and Services*, pp. 197–205, ICST, 2014.

[86] N. Y. Hammerla, J. Fisher, P. Andras, L. Rochester, R. Walker, and T. Pl{\"o}tz, "PD Disease State Assessment in Naturalistic Environments Using Deep Learning," in *AAAI Conference on Artificial Intelligence (AAAI)*, pp. 1742—-1748, AAAI Press, 2015.

[87] A. Sathyanarayana, S. Joty, L. Fernandez-Luque, F. Ofli, J. Srivastava, A. Elmagarmid, S. Taheri, and T. Arora, "Impact of Physical Activity on Sleep:A Deep Learning Based Exploration," *arXiv preprint arXiv:1607.07034*, jul 2016.

[88] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in *International Conference on Machine Learning (ICML)*, pp. 807–814, 2010.

[89] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[90] X. Glorot and A. Bordes, "Deep Sparse Rectifier Neural Networks," in *nternational Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 15, pp. 315–323, 2011.

[91] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors," *arxiv*, vol. abs/1207.0, 2012.

[92] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks From Overfitting.," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[93] L. Wan, M. Zeiler, S. Zhang, Y. L. Cun, and R. Fergus, "Regularization of Neural Networks Using DropConnect," in *International Conference on Machine Learning (ICML-13)*, pp. 1058–1066, 2013.

[94] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," *arXiv preprint arXiv:1503.02531*, 2015.

[95] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "Fitnets: Hints for Thin Deep Nets," in *International Conference on Learning Representation (ICLR)*, 2015.

[96] A. A. Rusu, S. G. Colmenarejo, Ç. Gülçehre, G. Desjardins, J. Kirkpatrick, R. Pascanu, V. Mnih, K. Kavukcuoglu, and R. Hadsell, "Policy Distillation," in *International Conference on Learning Representation (ICLR)*, 2016.

[97] J. Ba and R. Caruana, "Do Deep Nets Really Need to Be Deep?," in *Advances in neural information processing systems (NIPS)*, pp. 2654–2662, 2014.

[98] H. Sagha, S. T. Digumarti, J. d. R. Millán, R. Chavarriaga, A. Calatroni, D. Gen, and G. Tröster, "Benchmarking Classification Techniques Using the Opportunity Human Activity Dataset," in *INSS*, pp. 36–40, IEEE, 2011.

[99] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha, and Others, "Collecting Complex Activity Datasets in Highly Rich Networked Sensor Environments," in *Networked Sensing Systems (INSS), 2010 Seventh International Conference on*, pp. 233–240, IEEE, 2010.

[100] A. Stisen, H. Blunck, S. Bhattacharya, T. S. Prentow, M. B. Kjærgaard, A. Dey, T. Sonne, and M. M. Jensen, "Smart Devices Are Different: Assessing and Mitigatingmobile Sensing Heterogeneities for Activity Recognition," in *ACM Conference on Embedded Networked Sensor Systems*, pp. 127–140, ACM, 2015.

[101] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *International Conference on Learning Representation (ICLR)*, 2015.

[102] H. Cao, M. N. Nguyen, C. Phua, S. Krishnaswamy, and X. Li, "An Integrated Framework for Human Activity Classification.," in *International Conference on Ubiquitous Computing (UbiComp)*, pp. 331–340, 2012.

[103] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A Theory of Learning From Different Domains," *Machine learning*, vol. 79, no. 1-2, pp. 151–175, 2010.

[104] M. Long and J. Wang, "Learning Transferable Features With Deep Adaptation Networks," *arXiv preprint arXiv:1502.02791*, 2015.

[105] C. Louizos, K. Swersky, Y. Li, M. Welling, and R. Zemel, "The Variational Fair Auto Encoder," *arXiv preprint arXiv:1511.00830*, 2015.

[106] H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, and M. Marchand, "Domain-Adversarial Neural Networks," in *NIPS Workshop on Transfer and Multi-Task Learning: Theory meets Practice*, 2014.

[107] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, "Contractive Auto-Encoders: Explicit Invariance During Feature Extraction," in *International Conference on Machine Learning (ICML)*, pp. 833–840, 2011.

[108] M. Ghifary, W. Bastiaan Kleijn, M. Zhang, and D. Balduzzi, "Domain Generalization for Object Recognition With Multi-Task Autoencoders," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2551–2559, 2015.

[109] J. Schmidhuber, "Learning Factorial Codes by Predictability Minimization," *Neural Computation*, vol. 4, no. 6, pp. 863–879, 1992.

[110] Y. Ganin and V. Lempitsky, "Unsupervised Domain Adaptation by Backpropagation," *arXiv preprint arXiv:1409.7495*, 2014.

[111] H. Edwards and A. Storkey, "Censoring Representations With an Adversary," in *International Conference on Learning Representation (ICLR)*, 2016.

[112] M. Zhang and A. A. Sawchuk, "Usc-had: a daily activity dataset for ubiquitous activity recognition using wearable sensors," in *Proc. of UbiComp*, pp. 1036–1043, ACM, 2012.

[113] L. Sweeney, "Achieving K-Anonymity Privacy Protection Using Generalization and Suppression," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 571–588, 2002.

[114] P. Samarati, "Protecting Respondents Identities in Microdata Release," *IEEE transactions on Knowledge and Data Engineering*, vol. 13, no. 6, pp. 1010–1027, 2001.

[115] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-Diversity: Privacy Beyond K-Anonymity," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 1, no. 1, p. 3, 2007.

[116] N. Li, T. Li, and S. Venkatasubramanian, "T-Closeness: Privacy Beyond K-Anonymity and L-Diversity," in *IEEE International Conference on Data Engineering*, pp. 106–115, IEEE, 2007.

[117] D. Yu, K. Yao, H. Su, G. Li, and F. Seide, "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, IEEE, 2013.

[118] J. Xue, J. Li, D. Yu, M. Seltzer, and Y. Gong, "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, IEEE, 2014.

[119] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *arXiv preprint arXiv:1502.03167*, 2015.

[120] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-End Factor Analysis for Speaker Verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 788–798, 2011.

[121] G. Saon, H. Soltau, D. Nahamoo, and M. Picheny, "Speaker Adaptation of Neural Network Acoustic Models Using I-Vectors.," in *ASRU*, pp. 55–59, 2013.

[122] P. Karanasou, Y. Wang, M. J. F. Gales, and P. C. Woodland, "Adaptation of Deep Neural Network Acoustic Models Using Factorised I-Vectors.," in *INTERSPEECH*, pp. 2180–2184, 2014.

[123] K. Yao, D. Yu, F. Seide, H. Su, L. Deng, and Y. Gong, "Adaptation of Context-Dependent Deep Neural Networks for Automatic Speech Recognition," in *Spoken Language Technology Workshop (SLT), IEEE*, pp. 366–369, IEEE, 2012.

[124] S. M. Siniscalchi, J. Li, and C.-H. Lee, "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.

[125] D. Lopez-Paz, L. Bottou, B. Schölkopf, and V. Vapnik, "Unifying Distillation and Privileged Information," in *International Conference on Learning Representation (ICLR)*, 2016.

[126] O. Chapelle and A. Zien, "Semi-Supervised Classification by Low Density Separation.," in *AISTATS*, pp. 57–64, 2005.

[127] D.-H. Lee, "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.

[128] J. Weston, F. Ratle, H. Mobahi, and R. Collobert, "Deep Learning via Semi-Supervised Embedding," in *Neural Networks: Tricks of the Trade*, pp. 639–655, Springer, 2012.

[129] S. Rifai, Y. N. Dauphin, P. Vincent, Y. Bengio, and X. Muller, "The Manifold Tangent Classifier," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 2294–2302, 2011.

[130] Y. Grandvalet and Y. Bengio, "Semi-Supervised Learning by Entropy Minimization," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 529–536, 2004.

[131] P. Swietojanski and S. Renals, "Learning Hidden Unit Contributions for Unsupervised Speaker Adaptation of Neural Network Acoustic Models," in *Spoken Language Technology Workshop (SLT), IEEE*, pp. 171–176, IEEE, 2014.

[132] P. Swietojanski and S. Renals, "Differentiable Pooling for Unsupervised Speaker Adaptation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4305–4309, IEEE, 2015.

[133] S. Han, H. Mao, and W. J. Dally, "Deep Compression: Compressing Deep Neural Network With Pruning, Trained Quantization and Huffman Coding," in *International Conference on Learning Representation (ICLR)*, 2016.

[134] Z. Lin, M. Courbariaux, R. Memisevic, and Y. Bengio, "Neural Networks With Few Multiplications," *International Conference on Learning Representation (ICLR)*, 2016.

[135] M. Zhang and A. A. Sawchuk, "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.

[136] T. Huynh and B. Schiele, "Towards Less Supervision in Activity Recognition From Wearable Sensors," in *IEEE International Symposium on Wearable Computers (ISWC)*, pp. 3–10, IEEE, oct 2006.

[137] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network With a Local Denoising Criterion," *Journal of Machine Learning Research*, vol. 11, no. Dec, pp. 3371–3408, 2010.

[138] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *arxiv*, dec 2013.

[139] A. Rasmus, M. Berglund, M. Honkala, H. Valpola, and T. Raiko, "Semi-Supervised Learning With Ladder Networks," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 3546–3554, 2015.

[140] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved Techniques for Training GANs," *arxiv*, jun 2016.

[141] A. M. Khan, Y.-K. Young-Koo Lee, S. Y. Lee, and T.-S. Tae-Seong Kim, "A Triaxial Accelerometer-Based Physical-Activity Recognition via Augmented-Signal Features and a Hierarchical Recognizer," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, pp. 1166–1172, sep 2010.

[142] T. L. M. van Kasteren, G. Englebienne, and B. J. A. Kröse, "Hierarchical Activity Recognition Using Automatically Clustered Actions," in *International Joint Conference on Ambient Intelligence*, pp. 82–91, Springer Berlin Heidelberg, 2011.

[143] M. Laakso, T. Sarjakoski, and L. T. Sarjakoski, "Improving Accessibility Information in Pedestrian Maps and Databases," *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 46, no. 2, pp. 101–108, 2011.

[144] C. Cardonha, D. Gallo, P. Avegliano, R. Herrmann, F. Koch, and S. Borger, "A Crowdsourcing Platform for the Construction of Accessibility Maps," in *International Cross-Disciplinary Conference on Web Accessibility*, p. 26, ACM, 2013.

[145] T. Miura, K. Yabu, S. Ikematsu, and A. Kano, "Barrier-Free Walk: A Social Sharing Platform of Barrier-Free Information for Sensory/Physically-Impaired and Aged People," *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2927—-2932, 2012.

[146] K. Shigeno, S. Borger, D. Gallo, R. Herrmann, M. Molinaro, C. Cardonha, F. Koch, and P. Avegliano, "Citizen Sensing for Collaborative Construction of Accessibility Maps," in *International Cross-Disciplinary Conference on Web Accessibility*, W4A '13, (New York, NY, USA), pp. 24:1—-24:2, ACM, 2013.

[147] K. Hara, V. Le, and J. Froehlich, "Combining Crowdsourcing and Google Street View to Identify Street-Level Accessibility Problems," pp. 631–640, 2013.

[148] K. Hara, "Scalable Methods to Collect and Visualize Sidewalk Accessibility Data for People With Mobility Impairments," in *Annual ACM symposium on User interface software and technology*, pp. 1–4, ACM, 2014.

[149] N. Kern, B. Schiele, and A. Schmidt, "Multi-Sensor Activity Context Detection for Wearable Computing," in *European Symposium on Ambient Intelligence*, pp. 220–232, Springer Berlin Heidelberg, 2003.

[150] M. Verleysen, A. Ghio, L. Oneto, X. Parra Perez, and J. L. Reyes Ortiz, "A Public Domain Dataset for Human Activity Recognition Using Smartphones," in *International European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, p. 581, 2013.

[151] S. Dernbach, B. Das, N. C. Krishnan, B. L. Thomas, and D. J. Cook, "Simple and Complex Activity Recognition Through Smart Phones," in *International Conference on Intelligent Environments*, pp. 214–221, IEEE, jun 2012.

[152] Y. Hattori, S. Inoue, and G. Hirakawa, "A Large Scale Gathering System for Activity Data With Mobile Sensors," in *Annual International Symposium on Wearable Computers (ISWC)*, pp. 97–100, IEEE, jun 2011.

[153] T. van Kasteren, A. Noulas, G. Englebienne, and B. Kröse, "Accurate Activity Recognition in a Home Setting," in *International conference on Ubiquitous computing (UbiComp)*, (New York, New York, USA), pp. 1—-9, ACM Press, 2008.

[154] N. Kawaguchi, N. Nishio, N. Ogawa, Y. Iwasaki, K. Kaji, T. Terada, K. Murao, S. Inoue, Y. Kawahara, and Y. Sumi, "HASC Challenge: Gathering Large Scale Human Activity Corpus for the Real-World Activity Understandings," in *International Conference on Augmented Human (AH)*, (New York, New York, USA), pp. 1–5, ACM Press, 2011.

[155] O. Baños, M. Damas, H. Pomares, I. Rojas, M. A. Tóth, and O. Amft, "A Benchmark Dataset to Evaluate Sensor Displacement in Activity Recognition," in *ACM Conference on Ubiquitous Computing (UbiComp)*, (New York, New York, USA), pp. 1026—-1035, ACM Press, 2012.

[156] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas, "Window Size Impact in Human Activity Recognition," *Sensors*, vol. 14, no. 4, pp. 6474–6499, 2014.

[157] D. Figo, P. C. Diniz, D. R. Ferreira, and J. M. P. Cardoso, "Preprocessing Techniques for Context Recognition From Accelerometer Data," *Personal and Ubiquitous Computing*, vol. 14, no. 7, pp. 645–662, 2010.

付 録A　車いすセンシングによる路面アクセシビリティ情報収集システム

A.1　システム概要

ウェアラブルセンサによる行動認識を利用したアプリケーション事例として，車いすセンシングによる路面状況推定システムを構築した．図 A.1 に，提案システムの概要を示す．提案システムは，車いすユーザが屋外でとった行動やそのコンテキストを行動認識技術により推定し大量に蓄積することにより，路面状態を推定することを目的にしたものである．提案システムは，車いすユーザが持ちあるくスマートフォンにより車いすユーザの屋外での行動を計測し，計測されたセンサデータからアクセシビリティに関する情報を取り出すことでアクセシビリティの可視化や適切な経路選択，あるいは危険路面が近い場合に通知を行うシステムである．まず，スマートフォンに搭載されたセンシングアプリケーションにより車いすユーザの日常行動/周囲の環境情報を生データとして取得する．次に，分析サーバは計測された生のセンシングデータを受け取り，事前に学習したルール/モデルを利用して生のセンサデータからアクセシビリティに関する情報を推定する．推定された情報は地図上に可視化されたりより負担の少ないルートの選択に利用するなどしてアプリケーションとしてユーザに還元される．このように，スマートフォンにより計測可能なセンサデータを活用することでアクセシビリティ情報をより安価で大規模に蓄積することが本システムの目的である．

歩道の安全性は明らかに重要であるものの，都市の環境は車いすユーザ・視覚障害者などの歩行困難者にとって最適化されたものとはなっていない．都市のアクセシビリティ情報を可視化するアクセシビリティマップは，歩行困難者の移動に関わる負担を低減する主要な方法の 1 つであり [143]，データを効率的に集めるためのさまざまな研究がなされてきた．一方，現状のアクセシビリティ情報収集は，専門家

によるレビューないしクラウドソーシング [144, 145, 146, 147, 148] など人出に依存した方法に限られている．それゆえ，アクセシビリティ情報の共有は狭い場所に限定されているのが現状である．ウェアラブルセンサにより正確な車いす行動や走行環境に関するコンテキストの認識可能になれば，このようにこれまでは金銭的，時間的に困難であった実世界の観測を効率的に行えるようになる可能性がある．

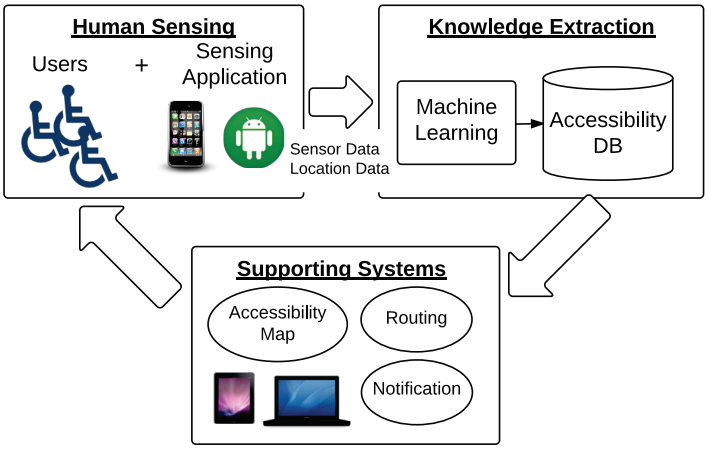


図 A.1: 車いすセンシングによるアクセシビリティ情報収集システムの概念図.

本論文がこれまで扱ってきた不特定ユーザに対する認識精度やプライバシーの観点は本システムにおいても重要な要素である．例えば車いすユーザがどこで転倒したかという情報は，路面が良いかどうかという情報と密接に絡んだ情報と考えられる．しかし，転倒行動を検出するためにアプリケーションを利用するユーザごとに転倒した場合のサンプルを得ることは難しい．このような，観測が困難，あるいは発生頻度が低い行動を検出するためユーザ依存認識ではなくユーザ独立認識が必要である．また，本システムはデータが大規模に得られる可能性があることが利点であり，そのためには限られたユーザではなく多くのユーザの行動を認識できることが重要である．このように大規模なユーザを対象にする場合，それぞれのユーザに対して教師付きデータを収集することは難しい．多くのユーザに参画してもらうためにはプライバシーに適切に配慮するなど，参入する心理的障壁を下げることも重要であろう．

本システムのように，行動認識を利用したシステムを実際に構築しようとする場合，学習データをどのように用意するかが問題となる．一方で，システム構築の立場からすると，学習データを作成することは時間的・金銭的なコストを伴う行為であり，可能な限り少なく済ませたい行為である．本章では，車いすセンシングによる路

面アクセシビリ情報収集システムの構築を題材に，利用するデータセットとディープネットの有効性の関係について検証し，ディープネットを利用する場合のデータセット収集指針について論じる．

ディープネットをはじめとして機械学習によるデータのモデル化を利用したアプリケーションを構築しようとする場合，学習データの取得方法には実験により新たにデータを取得する方法（実験アプローチ）の他に大別して既存の公開されたデータを利用する方法（転移アプローチ）が考えられる．転移アプローチのディープネットによるモデルを利用する場合における最も単純かつよく利用される方法はFine-Tuningと呼ばれる方法であり，何らかの行動認識タスク P を利用して学習したモデルのパラメタ θ_P を目的とする認識タスク P'（今回の例で言えば車いすセンシングデータから路面状態を表しうる行動を検出すること）に対して利用する方法である．この方法は，特に大量の一般データベースが公開されている画像認識や音声認識の領域では標準的に利用されている方法である．

行動認識においても，例えばHASCデータセットのように研究者間でデータセットを共有するプラットフォームを通して，2010年から2015年までの5年間で合計50,000以上の行動サンプルを収集したものや，あるいは本論文でも利用したOpportunityデータセットのようにより実際的な環境における27,000の行動を提供したデータセットなどが公開されている．ウェアラブルセンサを利用した行動認識用データセットの中で最も大規模なものでは，Inoueらが公開している看護行動センシングデータセット [8] は．合計で，22人の看護師に関するラベル付きデータセット（2週間分）を346.5時間分，また60人の看護師に関するラベル無しデータセット1655人日分を提供している．一方，詳細はA.2節で述べるが，転移アプローチが成功するためにはデータが大規模であるだけでなく参照するタスク P と解きたいタスク P' が何らかの意味で類似していることが必要であり，画像認識や音声認識などと比較して行動認識ではこの条件が満たされない場合が多い．そのため，行動認識をアプリケーションで利用する場合には多くの場合学習用のデータセットを自前で用意する必要がある．

A.3節以降では，車いすセンシングを利用した路面状態推定システムの構築事例を通して，行動認識によりこれまでは難しかった情報の収集が可能になることを示すとともに，ディープネットを利用する場合のデータ収集指針について論じる．具体的には，学習に利用するデータ数や学習に利用するユーザ数と認識精度の関係を検証することを通して，ディープネットをユーザ独立認識に利用する場合にどのようなデータセットを容易することが好ましいかについて考察する．また，A.2節で

は，既存の行動認識データセットについて整理し，行動認識を利用したアプリケーションにおいて転移アプローチを利用することの難しさについてまとめる．

A.2 ウェアラブルセンサを利用した行動認識に関するデータセットと転移アプローチの難しさ

最も古くからある公開されたデータセットの 1 つは，Kern らの研究によるものである [149]．Kern らは，自作した 3 軸加速度計則ノードをユーザの足首の上，膝の上，腰部分，手首，肘の上，の左右合計 12 箇所に設置し行動を計測している．計測時間は 18.7 分であり，計測人数は記載されていない．行動としては，座る，立つ，歩く，階段を登る，握手する，壁に書く，キーボードをうつの合計 8 種類が記録されている．

Verleysen らは，腰部分に設置した Galaxy S2 に搭載されている加速度センサとジャイロスコープ（50Hz）を利用して行動を計測している [150]．計測時間は合計 192 分であり，18 歳から 48 歳まで合計 30 人の行動を記録している．行動としては，立つ，座る，寝そべる，歩く，階段を登る，階段を下るの合計 6 種類が記録されている．

Dernbach らは，より複雑な行動を対象にしたデータセットとして，座る，立つ，歩く，走る，階段を昇り降りする，寝そべる，自転車に乗る，ドライブするといった比較的単純な行動に加えて洗濯する，料理する，治療を受ける（Medication），掃除する，手を洗う，植物に水をやるなどの複雑な行動を含む合計 14 種類の行動を対象にデータセットを作成している [151]．センサには Samsung Captive に搭載されている加速度センサとジャイロスコープを利用し，サンプリングレートはそれぞれ 30Hz としている．また，データ計測後にラベルをつける形ではなく行動をする前にユーザが行う行動を宣言し，それに付随してサンプルを計測する形をとっている．また，10 人の被験者による実験を行っている．

ウェアラブルセンサよりむしろ物体に設置されたセンサに強く焦点を当てた研究ではあるものの，Logan らは，家の中での行動計測用の大規模なデータセットとして，合計 104 時間にも及ぶデータを 900 個のセンサを利用して計測している．センサとしては，101 個のドアや棚，引き出しや窓等に設置された磁気センサ，36 個の温度計，10 個の湿度計，6 つの光度計，277 の物体に設置されたモーションセンサなどを利用している．被験者としては，1 組の夫婦のデータを計測している．行動

のラベルとして，コンピュータを利用している，音楽を聞いている，TV や映画を見ている，深く眠っている，本や新聞，雑誌を読んでいるなどの行動をラベル付けしている．

ウェアラブルセンサを利用した行動認識用データセットの中で最も大規模なものの 1 つに，Inoue らが公開している看護行動センシングデータセットがある [8]．センサとしては，胸ポケットに入れた iPod Touch に搭載された加速度センサと，腕に設置したリストバンド型の加速度センサを利用している．合計で，22 人の看護師に関するラベル付きデータセット（2 週間分）を 346.5 時間分，また 60 人の看護師に関するラベル無しデータセット 1655 人日分を取得している．ラベル付けは，ラベルデータ取得の対象となっていない看護師が観察者として，提供された iPod Touch 上のアプリケーションを利用してリアルタイムに付与している．行動クラスとしては事前に 41 の行動を準備しており，結果 25 クラスについて 5743 のラベルデータを取得している．

Hattori らは，13ヶ月にも及ぶ合計 200 人以上，35,000 以上の行動を記録した加速度データセット ALKAN を提供している [152]．行動の計測はミッションという単位で行われており，ユーザが行動前にモバイル端末上でどの行動をとるか宣言したうえでセンサにより行動を計測する形をとっている．216 個の iPod Touch を大学の学生及びにスタッフに貸出しており，1 日 1 回以上の行動入力を要請している．行動クラスとしては，日本の厚生労働省がまとめた”Exersize 2006”に従い，BBQ やダーツなどレクリエーション活動用の行動を追加したものを利用している．

Kasteren らは，1 人のユーザの合計 28 日間に及ぶ家庭内行動データを計測している [153]．行動のアノテーションにはユーザに設置した Bluetooth ヘッドセットによる音声データを利用し，ユーザが事前に行動を宣言する形をとっている．行動のクラスとしては，家にいない，トイレ中，シャワーを浴びている，朝食の準備をしている，夕食の準備をしている，飲み物の準備をしている，その他の合計 8 種類の行動を計測している．行動の数は合計 245 回となっている．

Kawahara らは，その他のデータセットが基本的に単一の研究グループにデータを収集するアプローチであるのに対して，行動データセットを共有するプラットフォームを通して，複数の研究者が独自に集めたデータを集約する形を提案している [154]．2010 年から 2015 年までの 5 年間で，合計 50,000 以上の行動サンプルが収集されている．

Banos らは，センサの設置位置のずれの影響の調査およびにセンサ位置のずれに頑健な手法の開発を目的に，9 個の慣性計測装置を利用した実験を行っている [155]．具

体的には，センサの設置箇所を 1)*Ideal-placement*：システム上望ましい位置にシステム提供者が設置した場合，2)*Self-placement*：システム提供者から指示を受けたユーザが自分でセンサの設置を行った場合，3)*Mutual-displacement*：故意に位置をシステム上望ましい場所からずらした場合，の 3 つを比較検討している．行動としては 33 個のフィットネス行動を対象としており，17 人からデータを取得している．センサには慣性計測装置 Xsens を利用し，左右の膝，左右の太もも，左右の手首，左右の上腕二頭筋，背中の合計 9 種類を設置している．*Ideal-placement* および に *Self-placement* について 15 時間ずつ，*Mutual-displacement* について 10 時間分のデータを取得している．

本論文での検証に利用したデータセットとして，Roggen らが提供する Opportunity データセットがある [99]．Opportunity データセットのという名前は，本データセットはその他のデータセットと違い明確な基準に従って行動を計測したり，あるいは行動計測後にシステム提供者がラベル付を行うのではなく，行動スクリプトを渡されたユーザが日和見的（Opportunistic）に行動を行うことに由来している．行動としては，移動状態レベル（立つ，座る，歩くなど 5 種類）とジェスチャーレベル（ドロワーを開ける，ドロワーを閉める，テーブルを片付ける，飲み物を飲むなど 18 種類）が提供されており，合計 27,000 の行動が計測されている．センサとしては，被験者の体に設置したセンサ（7 つの慣性計測装置，12 個の 3 軸加速度センサ）を含む 10 種類 72 個のセンサを利用している．合計 12 名の被験者 72 の試行のうち，4 名 24 の試行に対するラベル付きデータが提供されている．また，深層学習の有効性検証に標準的に利用されている [12, 19, 21]．

Zhang らが公開している USC-HAD（University of Southern California Human Activity Dataset) は，汎用的な低レベルな行動に関する大規模な行動データセット公開を試みてたデータセットの 1 つである．Opportunity や CMU-MMAC データセットが料理行動や家庭内行動のような特定領域の高レベル行動の認識を目的にしていたのに対して，USC-HAD はいずれのそれらの高レベル行動の背後にある基本的な行動 12 種類の行動の認識を目的としている．行動の種類としては，前進する，階段をのぼる，階段を下る，走る，ジャンプする，座る，立つ，眠る等の 12 種類である．

センサとしては 3 軸加速度センサと 3 軸ジャイロスコープの合計 6 軸のデータを計測しており，センサは被験者の腰付近に設置したポーチに装着されている．被験者としては，年齢や身長が多様な男女各 7 名ずつを対象に合計 14 名のデータを収集している．

ここまで説明したように，さまざまな研究者により行動データセットの収集と公

開は進められており，また例えば看護行動データセットのようにかなり大規模なデータセットも現れ始めている．一方，音声認識や画像認識などと比較して，これらのデータセットを実際にアプリケーション作成に利用することは難しい．その理由としては次のようなものがある．

センサに関する自由度

2 章でも述べたが，行動認識に利用されるセンサは多様であり，またそのセンサをどのように利用するかはアプリケーションにより異なる．例えば，Kern ら [149] や Dernbach ら [151] が加速度センサとジャイロスコープによる計測を行っているのに対して，Inoue らは加速度のみでの計測を行っている [8]．このように入力されるデータの種類自体が異なる場合にどのように転移アプローチを利用することは難しい．また，センサの種類だけでなくどの位置にどのように設置したかによってもセンサの観測値は大きく変化する．

観測ユーザに関する自由度とユーザ属性情報の制限

当然のことながら，データセットがどのユーザの行動を観測したものであるかはデータセットの計測者が設計するものであり自由度がある．本論文でもこれまで述べてきたように，同じ行動を取ったとしても観測される値はユーザにより異なる．すなわち，データセットはどのようなユーザの行動を計測したかによるユーザバイアスを含む．また，ユーザに関する精緻な属性情報は多くのデータセットにおいて公開されていない．

認識したい行動の多様性

1 章でも述べたが，認識したい行動はアプリケーションにより異なる．例えば，スポーツジムにおける行動をトラックする目的であれば腕立てや腹筋行動を検出する必要があるだろうし，あるいは医療目的でパーキンソン病の検出を行いたい場合にはパーキンソン病特有の歩き方を認識する必要がある．また，どのような行動を対象とするかにより利用するべき特徴は異なり，どのような情報を転移すればよいか，またあるタスクで獲得される特徴が別のタスクにおいても有用であるかを判断する

ことが難しい．これは，例えば音声認識が「あ」「い」「う」「え」「お」といった基本的な音素を識別するという点で共通であることと対照的である．

これらの理由から，行動認識における転移アプローチは難しい場合が多い．そのため，多くの応用研究では自前で用意したデータを利用しているのが現状である．

A.3　車いす屋外行動データセットの収集

システムのプロトタイプ作成にあたり，次の 4 つの観点に留意してセンシングシステムの設計や実験方法の選択を行い，車いす屋外行動データセットを取得した．

- 観点 1：実現可能性．ユーザの行動を邪魔しないセンサを利用すること．また，ユーザの行動を邪魔しない形でセンサを設置すること．
- 観点 2：有用性．多様な属性を持つユーザを含むこと．また，多様な路面状態を含むこと．
- 観点 3：自然さ．過度な制約をもうけずに，実環境に近い形のデータセットを作成すること．
- 観点 4：安全性．データ取得実験により，ユーザに過度の負担をかけないこと．

A.3.1　センシングシステム

図 A.2 にデータ取得実験の様子を示す．計測したデータは，行動認識の領域で最もよく利用されている [24]3 軸加速度センサ，位置情報，およびにアノテーション用の動画データの組み合わせである．図 A.2 にデータ取得実験の様子を示す．iPod touch を車いすの座席下に設置し，車いすユーザの行動を iPod touch に搭載された 3 軸加速度センサにより記録した．位置情報としては，準天頂衛星（Quasi Zenith Satellites System，QZSS）のレシーバを車いす後ろに取り付け，GPS 情報を取得した．実験の様子は車いす斜め後方からビデオカメラにより記録し，アノテーションデータの作成に利用した．加速度センサおよびにビデオカメラのサンプリング周波数はそれぞれ 50Hz およびに 30Hz とした．結果として，合計 1,255,900 の加速度サンプル（約 9 時間）を走行実験により取得した．

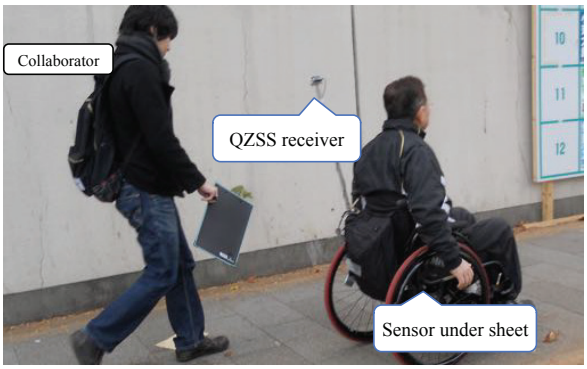


図 A.2: 実験に利用した車いす行動のセンシングシステムおよびに取得実験の様子.

A.3.2　実験参加者

データ取得実験（2013 年 11 月〜2013 年 12 月）には，日頃から車いすを利用している 9 名が参加した．9 名詳細は，7 名が男性と 2 名が女性，6 名が手動車いすで 3 名が電動車いす，また 20 歳から 60 歳と年齢にばらつきがあるユーザを選定した．

A.3.3　実験方法

計測環境の安全性や走行距離，路面の多様性などを考慮した上で東京四ツ谷駅近辺の経路を選定し，2013 年 11 月から 12 月にかけて行った．各参加者は，計測環境の安全性や走行距離，路面の多様性などを考慮した上で選定された上智大学周辺の決められた経路を 3 周走行した．走行経路および走行環境は図 A.3 に示す通りである．選定した経路はおよそ 1.5km であり，3 周でおおよそ 1 時間程度の走行時間であった．

実験により取得されるデータが可能なかぎり自然なものとなるよう，参加者に過度な負担をかけない範囲で多様な路面タイプを網羅するように心がけた（図 A.3 (b)-(g)）．また，走行環境や歩道上のどこを歩くかなどの制約をかけずに実験を行った．また，実験の安全性確保および経路の案内のために実験補助者が車いす後方を追尾する形で安全性を確認した．

A.3.4　アノテーション

分類対象として，ここでは路面アクセシビリティ情報を含みうる，1. 車道と歩道の間の縁石の乗り越え行動，2. 点字ブロック上の走行，3. 坂の登り降り，4. その他

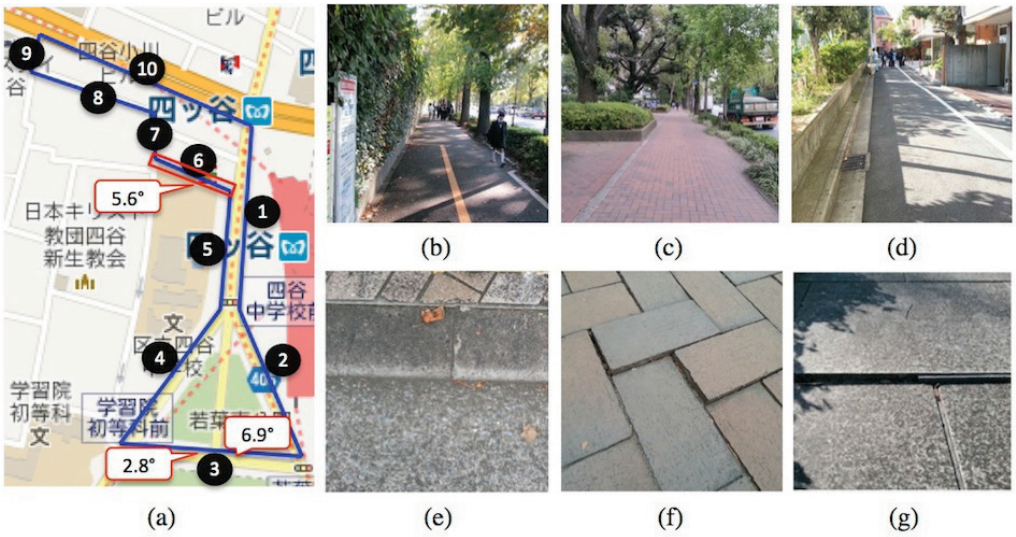


図 A.3: 走行経路に関する情報．(a) 経路概要, (b) 平らな路面, (c) タイル路面, (d) 坂（傾度 5 度以上）, (e) 車道と歩道の間の縁石, (f) タイル路面の間の小さな段差, (g) 木の根により盛り上がったコンクリート路面.

の路面の走行，5. 静止の 5 種類を選択し，動画データによるアノテーションを行った．具体的には，ビデオデータの各フレームを目視により上記の 5 つの行動のいずれに当たるかを確認記録した．最初の 4 つの行動は直接的に不快感およびに危険をともなう路面の存在を意味するものである．このような物理的な路面上の障害は車いす事故やインシデントを引き起こしうるものであるが，従来の人出によるアクセシビリティ情報収集では網羅が難しく，ディジタル化されていない重要な情報である．静止行動は背後にある路面アクセシビリティ障害を示しうる行動の一例である．例えば，交差点以外で車いすユーザがよく止まる場所などがあればそれは危険を示す可能性が高いと考えられる．

A.4 評価実験

A.4.1 実験 1：経験的特徴量を利用した基礎検討

実験概要

提案システムの実現可能性の検証にあたり，まず，特徴設計を含む従来的を利用したシステムによる基礎検討を行った．大筋の流れとしては一般的な行動認識のタ

スクと同様 [25]，1) 前処理，2) セグメンテーション，3) 特徴抽出，4）分類モデルの学習という流れで行った．

このプロセスにおける選択可能な要素として，前処理の方法，セグメンテーションサイズ，特徴量，分類器がある．評価にあたり，ここではそれらの要素を明示する．

まず，前処理としては移動平均フィルタによる平滑化を行った．これにより，ノイズ情報を低減するように務めた．ここで，フィルタ長としては実験的に 10 を選択した． セグメンテーションにおいては切り出すサブ時系列の長さ w およびに重なり率 p がパラメタとして存在する．ここで，重なり率 p とは k-1 番目のウィンドウ S_{k-1} と k 番目のウィンドウ S_k がどの程度重なっているかを意味するパラメタである．特に w の選択が分類精度に大きく寄与することが指摘されているものの [156]，その事前の選択は難しく実験的に決めるしかないのが現状である．ここでは、w として 50 から 1,000 まで変化させ認識精度の変化を確認することで，ウィンドウサイズの影響を調査した．同様の理由で，分類器の選択も重要であるものの事前の選択は難しい．ここでは，行動認識でよく利用される代表的な分類器である Support Vector Machine（SVM），ランダムフォレスト（RF），およびに K 近傍法（KNN）を利用してその認識精度を比較した．なお，SVM のカーネルには Radial Basis Function（RBF）を利用し，分類器が持つパラメタについてはそれぞれグリッドサーチにより検証した．

いくつかの研究が言及するように，特徴量の適切な設計は学習精度に大きく影響を与える重要なパラメタである [24, 157, 12]．しかし，適切な特徴量はタスクによって異なるため，ここでは行動認識に関する既存研究 [157, 12] を参考に次の 4 つの特徴量を検証した．

- **Raw Data**：本特徴量では特徴抽出を行わず，各セグメント S_k における 3 軸特徴量を $[x_k^T, y_k^T, z_k^T]$ のような形で単純に結合した．ここで x_k, y_k およびに z_k は k 番目のセグメントにおける各軸の加速度データに対応している．この方法により生成される特徴量はウィンドウサイズを w とした時に，$3w$ 次元である．

- **AveSTD**：AveSTD およびに次にの Heuristic では時間領域の特徴量を利用した．時間領域における特徴量は行動認識研究で最もよく利用される特徴量である．AveSTD では，スライディングウィンドウにより与えられた $3w$ $(w \times 3)$ の系列データを受けとり，x, y, z 各チャネルの平均値およびに標準偏差を算出した．この方法により生成される特徴量はウィンドウサイズによらず 6 次元である．

- **Heuristic**：Heuristic ではまず，スライディングウィンドウにより生成された系列データ x_k, y_k, z_k の各チャネルにおける階差数列 x_{kd}, y_{kd}, z_{kd} を計算した．その上で，$x_k, y_k, z_k, x_{kd}, y_{kd}, z_{kd}$ の各チャネルにおける平均値，標準偏差，最小値，およびに最大値を算出した．この方法により計測される特徴量はウィンドウサイズによらず 24 次元である．階差数列は高周波成分を強調した系列を生成するため，本特徴量は段差や点字ブロックなどの路面の粗さにより引き個々される加速度の突発的な変化に対して有効であると考えられる．

- **FFT**：FFT では，各チャネルの系列データを高速フーリエ変換により周波数領域へと変更した際に得られる係数を特徴量として利用した．FFT により得られる係数の利用は周波数領域における特徴量の利用時に最もよく利用される方法である．この方法により生成される特徴量はウィンドウサイズを w とした時に $3w$ 次元である．

分類結果

まずはじめに，図 A.4 に特徴量の認識精度に与える影響を調査した結果を示す．ここで，評価値として 10 交差検定を行った際の F 値を利用した．また，分類器としては SVM を利用した．平均 F 値により比較すると，各特徴量の認識精度はそれぞれ Raw Data（0.35），AveSTD（0.64），Heuristic（0.67），FFT（0.60）という結果であった．最も認識精度が高かった特徴量は Heuristic であり，4 つの行動をそれぞれ 0.63(1)，0.65(1)，0.50(2)，0.91(3) の精度で認識できていた．ここで，(1) とは各分類タスクにおける特徴量の順位を示している．Heuristic 特徴量を利用した分類器は特に縁石と点字ブロックの推定能力において他の特徴量を大きく上回っていることがわかる．最も認識精度が低かった特徴量は Raw Data であり，0.25 (4)，0.31 (4)，0.00 (4)，0.87(4) でありすべてのタスクにおいて最も認識精度が低かった．この結果からわかるもう一つの重要な結果は，Heuristic 特徴量がすべての問題に対してうまく行っているわけではない点である．例えば，坂の識別や静止行動の抽出においてはその他の手法と比較して認識精度の低下が見られた．この結果は，実務上特徴抽出器を分類したい対象の性質に合わせて設計することの重要性を示唆するものである．

次に，図 A.5 に，3 つの分類器の認識精度の比較結果を示す．ここで，特徴量には先程の実験で最も認識精度の高かった Heuristic を利用した．また，横軸は分類対象を，y 軸は SVM の認識精度を 0.00 とした場合の相対的な認識精度を示している．こ

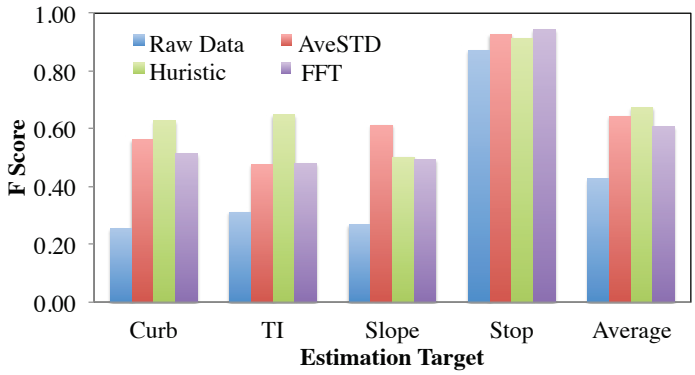


図 A.4: 特徴量と認識精度の関係の比較.

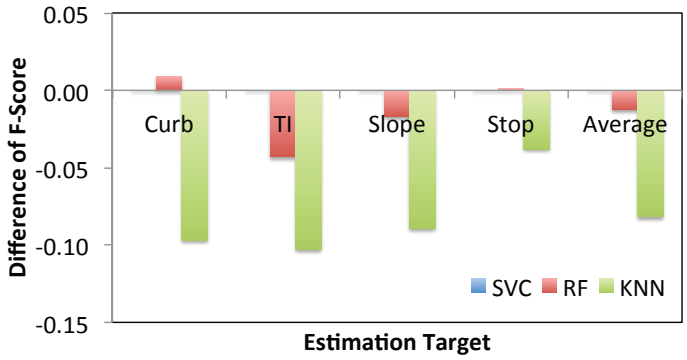


図 A.5: 分類器の比較.

の結果から，SVM がその他の分類器をほとんどの場合で上回っている事がわかる．特に，KNN と SVM の差は大きく，平均すると F 値で 0.08 ポイントの開きがあった．SVM と RF は特徴量の重要度を考慮する分類器であるが，KNN はすべての特徴量を同じ重みで取り扱う手法であり，この結果は特徴量の重要度合いが均等ではないことを示唆している．認識精度の改良には，特徴選択などの手法が寄与する可能性がある．

図 A.6 に，F 値と訓練に利用した学習データ数の関係を示す．ここで，x 軸は元の学習データ数を 1.0 とした時の相対的なデータ量であり，0.2，0.4，0.6，0.8，1.0 と変化させて実験した．教師あり学習によるモデル化を行う提案システムにおいて，センサデータへのアノテーション段階が最もコストのかかる部分であり，どの程度のラベル数があれば認識精度を確保できるのかは重要である．結果として，坂の分類以外においては訓練データ数の低下が認識精度にそこまで大きな影響を与えていないことがわかる．

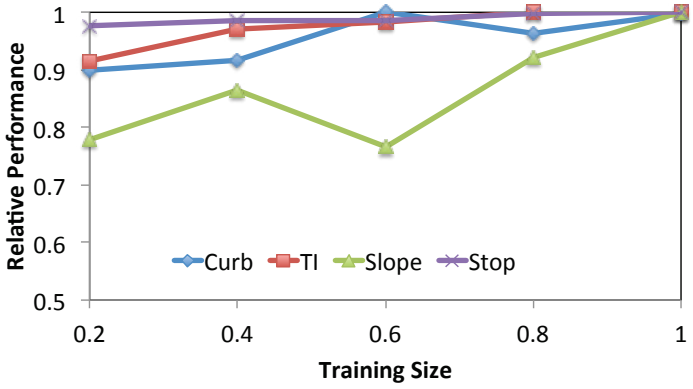


図 A.6: 訓練データ数と認識精度の関係.

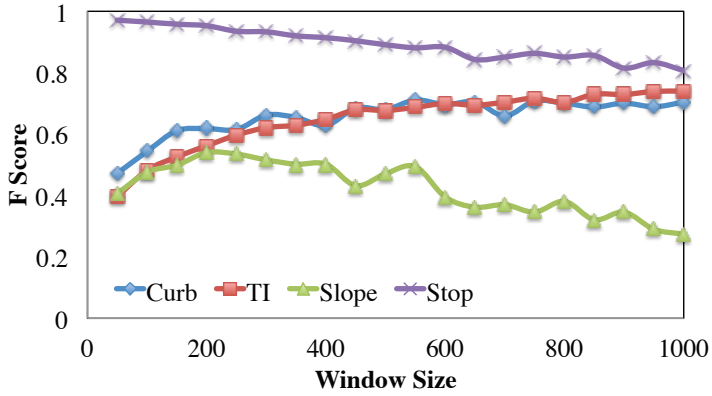


図 A.7: ウィンドウサイズがF値に与える影響.

図 A.7 には，F 値とスライディングウィンドウ時のウィンドウサイズ w の関係を示している．横軸がウィンドウサイズであり，50 から 1,000 まで 50 ずつ増加させた際の認識精度の変化を確認した．一般的な傾向として，縁石およびに点字ブロックについては徐々に認識精度が増加する傾向にあり，坂およびに静止行動の推定の場合には徐々に低下する傾向にあった．縁石乗り越え行動は 1.0 秒程度（50 フレーム）程度の行動であるにもかかわらず，長いウィンドウサイズがより高い認識精度となっていたのは興味深い結果である．大きなウィンドウサイズは各セグメントがより多くの情報を持つことを意味するため，段差や点字ブロックの推定タスクにおいてはその前後の加速度パターンも重要になることを示唆している．

図 A.8 に，推定された縁石（赤），点字ブロック（黄色），およびに坂（緑）を地図上に可視化した結果を示す．図中左側が推定結果であり，右側が実際の路面上対である．また，ウィンドウサイズは 400 に設定し，特徴抽出には Heuristic を利用し

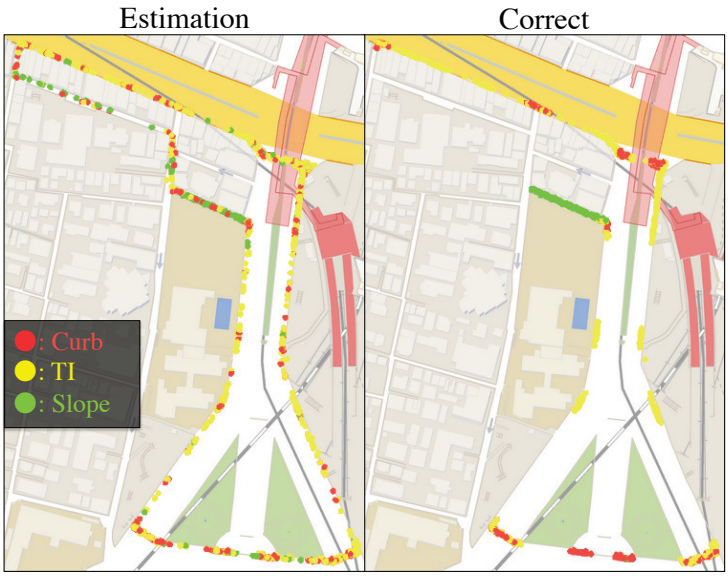


図 A.8: 27 周分の推定結果の可視化結果．縁石，点字ブロック，坂をそれぞれ赤，黄色，緑色の円で可視化した．

た．また，分類器としては SVM を利用した．なお，ここでは 27 周分合計 1,414 個の推定結果を Google Map API v3 を利用して単純に可視化した．可視化結果から，点字ブロックに関する精度に課題は残るものの，路面状態の大雑把な傾向が捉えられていることがわかる．本システムの特徴として同じ場所について大量のデータを取得しやすいという特徴があり，適切に処理することにより認識精度を更に高めることは可能であると考えられる．

A.4.2　実験 2：データセットと認識精度に関する検討

データセットと認識精度の関係を調査するにあたり，次の項目を評価した．なお，以降特に言及がない限り評価はユーザ独立認識の方法で行っており，検証データやテストデータという単語を利用した場合 4 章と同様の意味で利用するものとする．また，比較対象として利用している MV，Heu，および に ECDF は 4 章で利用したものと同様である．

- 学習に利用するユーザ数と認識精度の関係（表 A.1，図 A.9）
- 学習に利用するデータ数と認識精度の関係（図 A.10）

図 A.9 に，非個人モデルの学習に利用するユーザ数を変化させた際の認識精度の変化を可視化した結果を示す．横軸は学習に利用したユーザ数を表しており，縦軸が平均 F 値を意味している．図中の各丸は何らかの異なるユーザ群のデータにより学習された認識精度を示しており，例えば学習ユーザ数を 3 とした場合には，${}_8C_3 \times 9$ 通りの全組み合わせを検証している．また，図中の各線は，包含関係のあるデータセット間での認識精度の変化を示している．ここで包含関係とは，例えば訓練ユーザ数が 3 のときに利用したユーザが $\{P1, P2, M1\}$ であったときに訓練ユーザ数が 5 のときに利用したユーザが $\{P1, P2, M1, M2, M3\}$ や $\{P1, P2, M1, M2, M6\}$ のようにより少ない人数を利用した場合のユーザすべてを含むような場合である．すなわち，この直線の傾きが正の場合にはユーザ数の追加により認識精度が改善したことを，逆にこの直線の傾きが負の場合には認識精度が低下したことを意味する．また，表 A.1 に，ユーザ数を増やした場合に認識精度が改善していた割合を示す．例えば表 A.1-(a) の 1→3 の列において MV 行が 0.30 であるのは，学習ユーザ数を 1 から 3 に増やした場合に検証データに対する認識精度が改善したのは 30%であったことを意味する．

結果として，次のような結果が確認された．まず第 1 に，検証データについて，ユーザ数を増やした場合に認識精度が改善する傾向にあったのは CNN のみであった．その他の手法については，Heu でユーザ数を 7 に増やした場合をのぞいて改善率が 0.50 以下となっており，平均して認識精度が低下する傾向にあった．CNN を利用する場合でもその改善幅は小さくなる傾向が見て取れるものの，表 A.1 に示すように改善はユーザ数を増やしても継続的に続いていた．学習に利用するユーザを増やすことは，より大量のデータを利用可能にする一方で，入出力関係を複雑にし得る行為である．この結果は，複雑になった入出力関係を適切にモデル化するために

表 A.1: 学習に利用するユーザ数を増やした際に認識精度が改善したも割合の比較.

(a) 検証データ

	1→3	3→5	5→7	7→8
MV	0.30	0.48	0.39	0.43
Heu	0.36	0.50	0.59	0.33
ECDF	0.29	0.46	0.40	0.43
CNN	0.76	0.76	0.57	0.88

(b) テストデータ

	1→3	3→5	5→7	7→8
MV	0.66	0.60	0.54	0.4
Heu	0.71	0.61	0.56	0.43
ECDF	0.69	0.60	0.58	0.45
CNN	0.94	0.83	0.74	0.69

はニューラルネットのように表現力の高いモデルを利用する必要があることを示唆している.

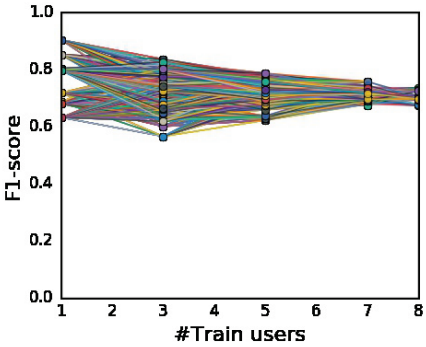
第 2 に，テストデータに対する認識精度の改善度合いでも CNN がその他の手法と比較して認識精度を改善する傾向が強かった．その他の手法についても，表 A.1-b に示すように 7→8 を除けば平均すると認識精度は改善していたが，その改善度合いは最大でも 0.71（Heu，1→3）であり，残りの 30% 程度は認識精度が低下していた．一方 CNN は 1→3 においては 94%が認識精度改善しており，またユーザ数が増えた場合でも認識精度は平均すると改善していた.

次に，図 A.10 に，学習に利用するデータ量を減らした際のテストデータに対する認識精度の変化を示す．ここで各横軸は 8 人全データを利用した場合のデータ量を 1 とした際の割合であり，縦軸は平均 F 値である．図 A.10-a が利用するユーザを減らした場合の結果であり，A.10-b はユーザ数は 8 に固定したまま 1 人あたりの利用データ数を減らした結果である．また，図 A.10-c は a と b を比較したものであり，実線が a の結果に，点線が b の結果にそれぞれ対応している．図中各マーカおよびに色はそれぞれ特徴量の違いを表している.

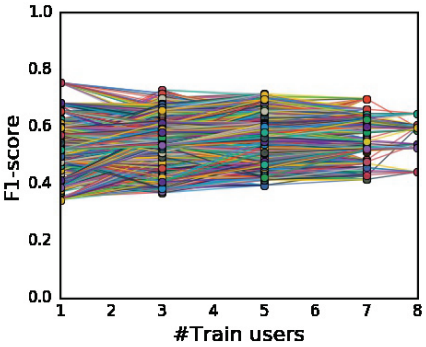
結果として，次のような結果が得られた．まず，データセットをすべて利用する場合には CNN の認識精度が顕著に高かった．また，例えば (b) の結果では MV，Heu，ECDF の 3 種類ではデータ数を 0.5 まで絞った場合と 1.0 の場合音認識精度がほぼ同じであり，データ量の増加による認識精度の改善には限界があることがわかる．次に，利用データ量を減らしていった場合を確認すると，学習ユーザを減らす場合にはおおよそ 0.3，また学習データを減らす場合にはおおよそ 0.2 程度で認識精度が初めて CNN が ECDF を下回っていた．今回のデータセットは 4 章でも述べたように合計 7,000 サンプル程度の比較的小規模なデータであり，またうち 1 人分をテスト

データ，残りの 10%を検証データに利用しているため学習に利用しているデータは平均するとおおよそ 7,000$\times\frac{8}{9}\times 0.9$ =5,500 程度である．必要なデータ数は学習データ数に依存するものであるため一般化は難しいものの，少なくとも今回のタスクにおける結果はデータが 1,000 程度の場合でも CNN を利用することで認識精度は改善する可能性があることを意味している．

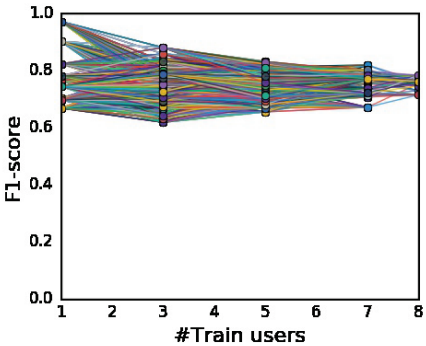
また，(c) の結果は，いずれの特徴量を利用する場合でも点線（ユーザ 1 人あたりのデータを減らす場合）と比較して実線（ユーザ数を減らす場合）の認識精度低下が大きいことを示していた．また，とくに学習手法として CNN を利用する場合にはユーザ数を減らした際の認識精度の劣化が顕著であった．この結果は，特に CNN を利用する場合に同じデータ量を用意するのであれば 1 人あたりデータ量を増やすよりもデータを収集するユーザ数を増やすことが重要であることを示唆している．



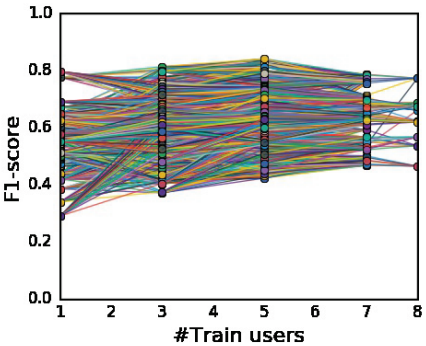
(a) MV 検証



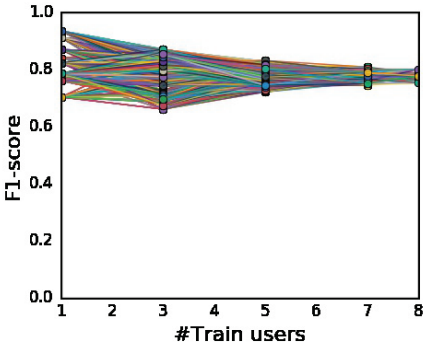
(b) MV テスト



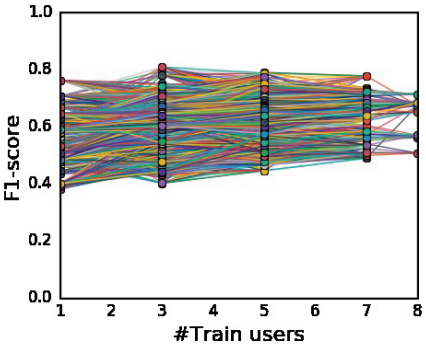
(c) Heu 検証



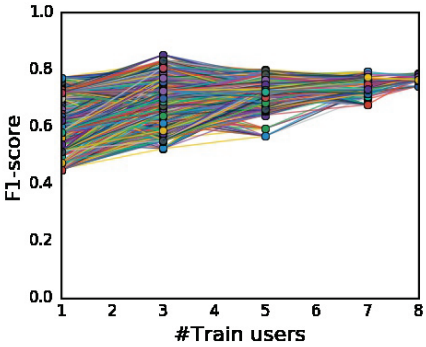
(d) Heu テスト



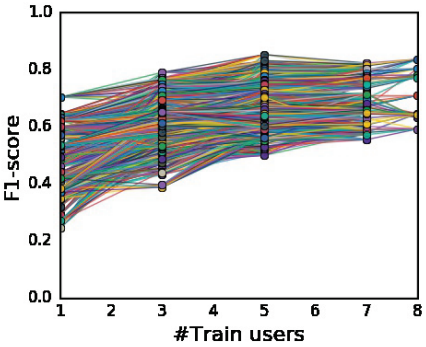
(e) ECDF 検証



(f) ECDF テスト

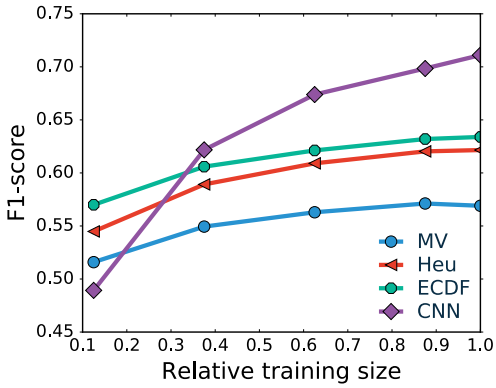


(g) CNN 検証

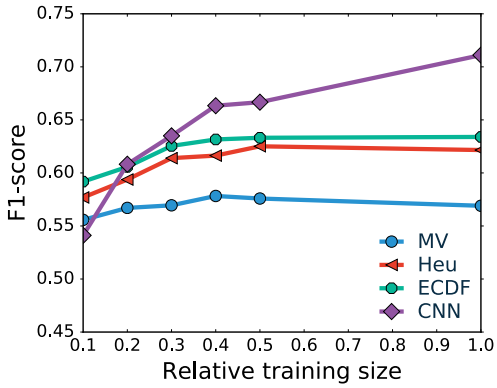


(h) CNN テスト

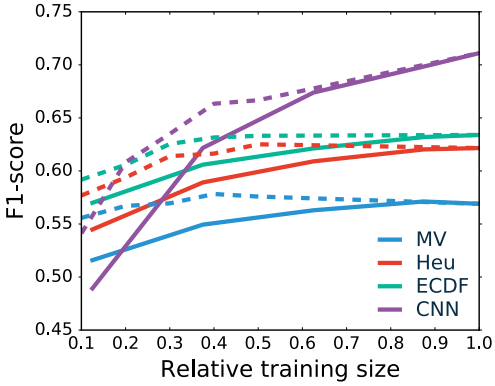
図 A.9: 非個人モデルの学習に利用するユーザ数を増加させた際の認識精度の変化．各行がそれぞれ各特徴量に，また左列が検証データに対する認識精度，右列がテストデータに対する認識精度を示している．



(a) 学習に利用するユーザ数を減らした場合



(b) 学習に利用するデータサイズを減らした場合



(c) a と b の比較

図 A.10: 学習に利用するデータ量に対する認識精度の関係.

付 録B　5章に関する付録

B.1　交差エントロピー最大化と相互情報量

$$
\begin{aligned}
I(U;R) &= \sum_{(u,r)} P(U=u,R=r)\log\frac{P(U=u,R=r)}{P(U=u)P(R=r)} \\
&= \sum_{n}^{N} \log\frac{P(U=u_n,R=r_n)}{P(U=u_n)P(R=r_n)} \\
&= \sum_{n}^{N} \log\frac{P(U=u_n|R=r_n)P(R=r_n)}{P(U=u_n)P(R=r_n)}
\end{aligned}
$$

ここで，パラメタ θ_u によるニューラルネットワークの入出力関係 $f_{u_n}(r_n;\theta_u)$ を事後確率 $P(U=u_n|R=r_n)$ のモデルとすると，

$$
\begin{aligned}
I(U;R) &= \sum_{n}^{N} \log\frac{f_{u_n}(r_n;\theta_u)}{P(U=u_n)} \\
&= \sum_{n}^{N} \log f_{u_n}(r_n;\theta_u) - \sum_{n}^{N} \log P(U=u_n)
\end{aligned}
$$

となる．この時，第2項目はデータが与えられた上では固定であるから，相互情報量 $I(U;R)$ を最大化することは事後確率の尤度 $\sum_n^N \log f_{u_n}(r_n;\theta_u)$ を最大化することに等しい．

また，式5.3の u_nk は，各サンプルがk番目のクラスに属するかを示すいずれかが1になるような値であるので，$\sum_n^N \log f_{u_n}(r_n;\theta_u)$=$N\cdot \mathrm{L}_u$ となる．

B.2 複数のウィンドウを利用した場合のユーザ予測精度に関する検討

B.1 に，複数のウィンドウを予測に利用した場合の結果を示す．ここで，複数のウィンドウを予測に利用する方法としては，単純に複数のウィンドウの出力 $P(u|r_i)(i \in 1 \cdots w)$ の平均値を取るアンサンブル処理を利用した．横軸は予測に利用したウィンドウ数 w を，縦軸はユーザ予測精度をそれぞれ表している．データセットとしては，5.2 の結果で MLP_{800} を γ として利用した場合に最も CNN と提案手法の差が大きかった USC を利用した．また，他ユーザでも概ね同じ傾向であったため，ここではユーザ ID=1 をテストデータとする Leave-One-Subject-Out イテレーションの結果を示す．結果として，CNN，$\mathrm{UANN}_{\mathrm{MLPA}}$ どちらもウィンドウを大きくすると w=1 の場合と比較して予測精度が高まる傾向にあった．ただし，例えば w=100 の場合には CNN と $\mathrm{UANN}_{\mathrm{MLPA}}$ を比較するとほぼ同程度であるが，CNN は w=10 の時点ですでにほぼ 1.0 であり，依然として提案手法により予測がわずかながら抑制されていることには注意されたい．

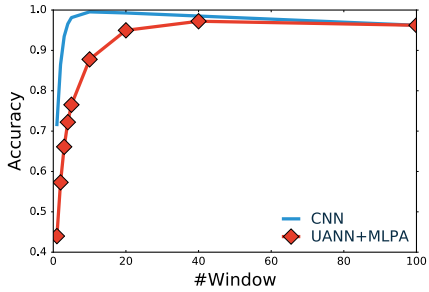


図 B.1: 複数のウィンドウでの出力確率の平均によるユーザ識別性能の変化．

付 録C　発表業績一覧

論文誌

[1] Yusuke Iwasawa, Ikuko Yairi, Yutaka Matsuo, ”Combining Human Action Sensing of Wheelchair Users and Machine Learning for Autonomous Accessibility Data Collection”, IEICE Transactions, 2016

[2] 岩澤有祐，矢入郁子，松尾豊，”深層学習による不特定ユーザ行動認識のための表現の学習とその分析”，電子情報通信学会論文誌，2016（投稿中）

[3] 岩澤有祐，矢入郁子，松尾豊，”ユーザ敵対型ニューラルネット– ウェアラブルセンサを利用した行動認識におけるユーザ汎化とプライバシーへの配慮のためのユーザ独立な表現の学習手法 –”，人工知能学会論文誌，2016（照会）

[4] 岩澤有祐，矢入郁子，松尾豊，”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”，人工知能学会論文誌，2016（照会）

[5] 長峯洸弥，岩澤有祐，松尾豊，矢入郁子，”スマートフォンを用いた車いす走行における身体的負担の評価”，電子情報通信学会論文誌，2016（照会）

国際学会発表

[1] Yusuke Iwasawa, Yutaka Matsuo, and Ikuko Eguchi Yairi, ”Toward an Automatic Road Accessibility Information Collecting and Sharing Based on Human Behavior Sensing Technologies of Wheelchair Users”, The 6th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN), Sep. 27-30, Berlin, Germany, 2015

[2] Yusuke Iwasawa, Koya Nagamine, Yutaka Matsuo, Ikuko Eguchi Yairi, “Road Sensing: Personal Sensing and Machine Learning for Development of Large

Scale Accessibility Map”, International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS, Poster), Lisbon, Portgul, 2016

[3] Koya Nagamine, Yusuke Iwasawa, Yutaka Matsuo, and Ikuko Eguchi Yairi, ”An estimation of wheelchair user’s muscle fatigue by accelerometers on smart devices”, AACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (UbiComp, Poster), Osaka, Japan, 2015

[4] Yusuke Iwasawa, and Ikuko Eguchi Yairi, ”Spatiotemporal Life-Log Mining of Wheelchair Users’ Driving for Visualizing Accessibility of Roads”, IEEE International Conference on Data Mining Workshops (ICDMW), Dec. 7-10, Dallas, TX, America, 2013

[5] Yusuke Iwasawa, Hidetaka Suzuki, and Ikuko Eguchi Yairi, “Detecting Exceptional Actions Using Wearable Sensors’ Data for Developing Life-Log Database of Visually Impaired People”, Association for the Advancement of Artificial Intelligence (AAAI) Spring Symposium, Mar. 24-26, Stanford, America, 2013

[6] Yusuke Iwasawa and Ikuko Eguchi Yairi, “Life-Logging of Wheelchair Driving on Web Maps for Visualizing Potential Accidents and Incidents”, Pacific Rim International Conference on Artificial Intelligence (PRICAI), Sep. 3-7, Kuching, Malaysia, 2012

[7] Ikuko Eguchi Yairi, Kumi Naoe, Yusuke Iwasawa, and Yusuke Fukushima, “Do Multi-touch Screens Help Visually Impaired People to Recognize Graphics?”, ACM SIGACCESS Conference on Computers and Accessibility (ASSETS, Poster), pp. 237-238, Oct. 24-26, Dundee, Scotland, 2011

解説記事

[1] 中山浩太郎，岩澤有祐，黒滝紘生，松尾豊，”Deep Learning の基礎と実装”，情報処理学会誌，2015

国内学会発表

[1] 高橋宏紀，岩澤有祐，長峯洸弥，松尾豊，矢入郁子，車いすセンシング加速度データによる路面特徴の数値化，ヒューマンインタフェースシンポジウム 2016，Sep. 2016

[2] 大森涼，高橋宏紀，長峯洸弥，岩澤有祐，松尾豊，矢入郁子，車いすセンシングによる走行時負担度の定量化とその評価，ヒューマンインタフェースシンポジウム 2016，Sep. 2016

[3] 岩澤有祐，矢入郁子，松尾豊，車いす振動レベルの蓄積による都市のアクセシビリティ推定，ユビキタスコンピューティング研究会，May. 2016

[4] 岩澤有祐，矢入郁子，松尾豊，深層行動認識モデルのユーザ特化圧縮，人工知能学会全国大会 2016，Jun. 9, 2016

[5] 長峯洸弥，落合真季，岩澤有祐，松尾豊，矢入郁子，スマートフォンを利用した手動車いす利用者の負担推定・共有システム，人工知能学会全国大会 2016，Jun. 9, 2016

[6] 岩澤有祐，矢入郁子，松尾豊，車いす行動センシング加速度データへの表現学習の適用，人工知能学会全国大会 2015，May. 31–Jun. 2015

[7] 長峯洸弥， 岩澤有祐， 松尾豊， 矢入郁子, スマートデバイスの三軸加速度センサを利用した車いす走行者の疲労推定, 人工知能学会全国大会 2015，May. 31–Jun. 2015

[8] 岩澤有祐，矢入郁子，多次元時系列データ解析によるアクセシビリティ可視化システムの開発 ，人工知能学会全国大会 2014，May. 12–14, 2014

[9] 長峯 洸弥, 岩澤 有祐, 矢入 郁子, スマートデバイスを用いた車いすライフログにおける転倒行動分析, ヒューマンインタフェースシンポジウム,2014

[10] 田口健人，岩澤有祐，矢入郁子，スマートデバイスの 3 軸加速度センサを用いた重度視覚障害者歩行分析の研究，ヒューマンインタフェースシンポジウム 2013

[11] 高田佑平，岩澤有祐，矢入郁子，3 軸加速度センサを用いた機械学習による車いす走行危険箇所可視化システム，ヒューマンインタフェースシンポジウム 2013，

[12] 岩澤有祐，矢入郁子，車いす走行ライフログの時空間解析による路面状況推定システム，人工知能学会全国大会 2013，Jun．4-7, 2013

[13] 塩谷駿介, 鈴木秀孝, 岩澤有祐, 矢入郁子, 秋葉原駅前歩行実験による重度視覚障害者の心理・歩行特性分析，ヒューマンインタフェースシンポジウム 2012, pp. 905-912, Sep. 4-7, 2012

[14] 高橋雄介 , 浅井信尊, 岩澤有祐, 矢入郁子, 三軸加速度センサを用いた車いすバスケットボール選手の動作分析，ヒューマンインタフェースシンポジウム 2012, pp.905-912, Sep. 4-7, 2012

[15] 岩澤有祐，矢入郁子，3 軸加速度時系列データからの車いす走行行動分析の研究，人工知能学会全国大会 2012, Jun. 12-15, 2012

[16] 尚永久美，岩澤有祐，福島裕介，矢入郁子，マルチタッチディスプレイを用いた視覚障害者への図形提示インタフェースとその評価，ヒューマンインタフェースシンポジウム 2011, Sep. 13-16. 2011.

謝辞

本博士論文は，筆者が東京大学大学院工学系研究科技術経営戦略学専攻に在学中に行った研究をまとめたものです．本研究を行うにあたり，非常に多くの方々のご指導ご鞭撻を賜りましたことを，この場を借りて心より感謝いたします．

指導教官である松尾豊准教授には，非常にお忙しい中貴重な時間を割いてご指導頂きました．直接的に研究内容についてご指導いただいたことが直接的に本研究の内容に現れているだけでなく，研究者，指導者，そしてときに啓蒙者として精力的に学内外で活動される松尾先生のその姿勢から非常に多くのことを学んだことが血肉となり本博士論文を提出に至る推進力となりました．これまでのご指導ご鞭撻に心より感謝いたします．

本研究が対象として扱った深層学習の領域はハードウェアリソースが大量に必要になるものであり，松尾研究室の優れたハードウェアリソースなくして研究を円滑に進めることはできませんでした．私が松尾研究室に入る前から今に至るまで松尾研の研究環境を継続的に改良してきてくださった東京大学中山講師をはじめ，現博士 1 年の黒滝 紘生くん，現修士 1 年の味曽野 雅史くん，学術支援職員の Emilio Castillo さんなど，松尾研究室の優秀な HPC チームに感謝いたします．

松尾研究室の中山講師には，研究を進める上で悩んだ際に研究の進め方にアドバイスをいただくなど，研究をサポートしていただきました．さまざまな意味で露出度が高まっていく松尾研を運営するという，非常にハードな仕事をこなす身でありながら，親身に相談にのっていただけたことが博士論文執筆を進める上での支えの 1 つとなりました．心より感謝います．

本研究の成果の一部は筆者が上智大学理工学研究科情報学専攻に在籍していた修士時代の研究成果を発展させたものであり，修士時代の恩師である矢入郁子准教授には博士課程への進学からメンタルケアに至るまで研究活動全般に渡る多大なご支援を賜りました．学部 4 年のときに矢入研究室に入った際にはすでに就職先を決めて就職する予定だった私がこうして博士論文を書くまでになったことは自分で思い

返しても不思議ですが，修士時代に矢入研究室に入り矢入先生のご指導がなければそもそも博士課程に進学することもなかったのではないかと思います．矢入先生への感謝は記しきれないほどありますが，今後は少しでもかけていただいた御恩に報いれるよう邁進してまいります．

本博士論文の付録に記載しました車いすデータセットの収集およびにアクセシビリティ情報収集システムの研究開発には，矢入先生だけでなく上智大学の先輩・同輩・後輩などさまざまな方のご協力をいただきました．特に，現矢入研究室修士 2 年の長峯 洸弥くん，修士 1 年の大森 涼くん，学士 4 年の高橋 宏紀くん，卒業生の岡田　遼太郎くん，植松 洋介くん，植松 啓誠さんにはデータの収集やデータ収集用アプリケーションの開発，データ収集実験の補助などさまざまな面でサポート頂きました．また，車いす走行実験を行うにあたってはエリアワークス株式会社様の河野様，明珍様には多大なるご協力を頂きました．特に，河野様は毎回の実験に参加くださり，被験者の対応など後ろ側から実験を協力にサポートしてくださいました．また，実験に参加してくださった車いす被験者の方々のおかげで本研究は実現することが出来ました．毎回 6km 程度走行するという，特に手動車いすの方には体力的にもなかなか大変な実験にも関わらず笑顔で，パワフルに取り組んでくださる姿をみて，よりいっそう自身の研究に取り組む契機となりました．心より感謝いたします．

松尾研究室秘書の永本 登代子さん，浪岡 亮子さん，また 2016 年 12 月 1 日に退職された中野 佐恵子さんには，事務的な書類作成などについて日々サポート頂きました．秘書の方々をはじめとして，当たり前のことを当たり前にこなしてくださるがいるからこそ，学生を初め研究室のメンバーが研究という自分の興味に向かって邁進するということが許されているのだと改めて感じました．心より感謝しています．

松尾研究室の先輩・後輩・そして同期の野中 尚輝くんには，研究への助言やサポート等，様々な面でお世話になりました．研究だけでなく，皆様との研究生活の中で，さまざまな刺激，自分にはなかった知識や考え方を得ることが出来ました．研究室生活におけるすべての経験が，本研究のバックグラウンドとなっています．苦楽をともにした皆様に感謝の意を示します．

最後になりますが，博士課程というおそらく何のことかよくわからなかったであろう道に進む私を，暖かく見守ってくれた家族全員に感謝します．

なお，本研究は日本学術振興会特別研究員奨励費 (DC1，26-5331) の助成よって

行われました.