

端末所持位置推定における転移学習の有効性検証

桂田 連¹ 長谷川 達人¹

概要：本研究では、畳み込みによってネットワーク内部で自動的に特徴量を抽出する Convolutional Neural Network (CNN) を用いた所持位置推定モデルを開発する。また、本研究で用いる所持位置ラベル付き歩行データセットは基本行動認識分野で公開されているようなデータセットと比べ小規模である。この性質が CNN の性能に影響を及ぼす可能性があるため、本研究では大規模データセットが公開されている行動認識モデルから、所持位置推定モデルへ転移学習する手法も提案し、有効性を確認する。所持位置推定の先行研究では、予め有効と考えられる特徴を人間の知識に基づいて設計していたが、CNN を用いることで、新たに有効な特徴表現を自動抽出し、推定精度向上が期待できる。評価実験の結果、独自の CNN モデルを開発したが、転移学習なしで学習を進めた場合は先行研究の性能には及ばなかった。しかし、転移学習を導入することにより被験者全員で転移前より約 10% 推定精度を向上させることができ、先行研究より約 5% の精度向上が確認された。被験者によっては転移前後で推定精度が 10% 以上推定精度が向上する者もあり、本実験によって、転移学習を導入した CNN の有効性を示した。また、転移学習手法にも全体的なモデルの転移や、部分的な転移等複数の手法が考えられる。そこで、本研究では転移させる層について様々な転移手法を評価することで、モデルの内半分ほど転移させる手法が有効であることも明らかにし、所持位置推定の精度向上に貢献する。

Effectiveness Verification of Transfer Learning in Sensor Device Position Estimation

REN KATSURADA¹ TATSUHITO HASEGAWA¹

1. はじめに

本研究では、スマートフォンの所持位置推定という課題に対して、CNN を利用した推定モデルを開発する。CNN による畳み込みとプーリングにより特徴表現を自動的に学習する為、手動で特徴を抽出する場合とは異なった結果が期待される。ただし、CNN のような深層学習は、大量のデータセットを必要とする性質上、本研究で取得した比較的小規模データセットでは不向きであることが懸念される。本研究ではその対策案として CNN による学習済み行動認識モデルを利用した転移学習を応用し、その有効性を検証する。また、転移学習手法としても、中間層を部分的に転移する手法や、モデル全体を転移させる手法のように複数転移手法が考えられるが、本研究ではそのような転移部分に着目した手法の中から、所持位置推定に対して有効

な手法を明らかにする。本研究における転移学習の概要を図 1 に示す。図 1 のように、本研究では行動推定を Source Domain (SD)、所持位置推定を Target Domain (TD) とし

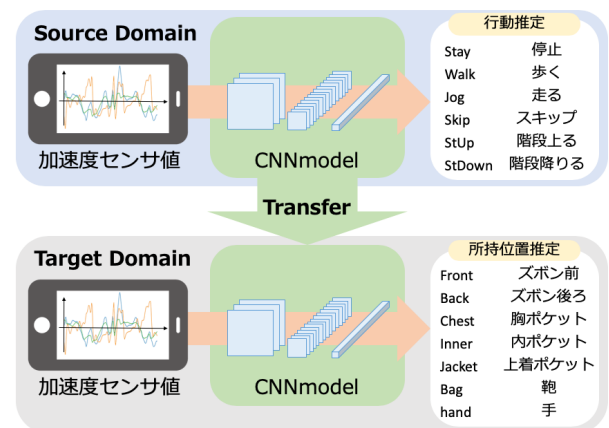


図 1: 本研究における転移学習の概要

¹ 福井大学大学院工学研究科情報メディア工学専攻
Graduate School of Engineering, University of Fukui

て転移学習を行う。まず SD について CNN モデルを学習させ、重み情報を保存しておく。次に TD について、保存した重みを用いて再学習させる手法を本研究における転移学習手法とする。

本研究における行動認識とは、HASC コーパス [1] にて記録されている基本行動 6 種類 (Stay (停止), Walk (歩く), Jog (走る), Skip (スキップ), StUp (階段上る), StDown (階段降りる)) を加速度センサ等の値から分類する課題である。

対して、所持位置とは、位置の座標情報ではなく、スマートフォンが所持されている位置 (ズボンポケット, 手, 鞆など) の名称を指す。本研究の所持位置推定ではポケット (ズボン前, 後ろ, ジャケット, 内, 胸), 手, 鞆の計 7 種類を推定対象とする。所持位置推定は、歩く走る停止するという行動の中でも、さらに各行動に対して各所持位置を推定する課題であるため、分類難易度が高い推定課題であると言える。所持位置推定を実現できると、位置に応じてスマートフォンの通知音量の自動調節や通知手段の変更、ポケットの中の誤作動の防止といったコンシューマサポートを実現できる。また、スマートフォンを用いて利用者の状況及びその周辺環境を推定する研究は近年注目を集めているが、センサから観測される値はセンサの所持位置に依存することが知られている [2]。よって、端末所持位置を正確に把握する技術の実現は行動認識性能の向上にも繋がる可能性がある。

転移学習について述べる。本研究における転移学習手法は、SD で訓練したモデルの重みを TD で訓練するモデルの初期値として再学習することを指す。画像認識の分野では、ILSVRC という画像のパターン認識モデルの性能を競う大会で用いられている大規模データセット ImageNet [3] が公開されている。ImageNet を用いて学習した CNN モデルが抽出した特徴量は汎用性の高いものであることが示されている [4]。また、同調査において、その特徴量を用いた転移学習は他の画像データセットにおけるパターン認識や、さらに高度な認識課題に対して、転移学習を行わない場合よりも良い性能を示すことが明らかになっている。しかし、筆者の調査した限りでは行動認識モデルの転移学習による効果、特に所持位置推定への転移学習の有効性については明らかにはなっていない。以上を踏まえた本研究による主要な貢献を以下に述べる。

- (1) 端末所持位置を推定する CNN モデルの開発
- (2) 行動認識モデルから所持位置推定への転移学習による推定精度変化の考察
- (3) 行動認識によって得た特徴量を所持位置推定に転移する場合に、どの層を、どの範囲で転移させると最も効果的であるかに関する考察

本稿の残りの構成は次の通りである。2 章で関連研究を踏まえた本研究の立ち位置を示し、3 章では提案手法を述

べる。4 章では提案手法を評価する実験概要とその結果、考察を述べ、5 章で本稿をまとめる。

2. 関連研究と本研究の立ち位置

関連研究として、本章ではまずスマートフォンを用いた行動認識に関する研究をいくつか挙げ、次に所持位置推定に関する研究を挙げる。その後に転移学習に関する研究を挙げた上で、本研究の立ち位置を明確にする。

行動認識は多くの研究が進められており、近年ではスマートフォンを用いた研究が盛んである。Anguita ら [5] はスマートフォンに内蔵されている加速度センサ等を用いて「直立」、「歩行」、「横になる」、「階段を上る」、「階段を下がる」といった基本的な動作を認識する手法を提案している。このような動作以外にも、Abbate ら [6] はスマートフォンに内蔵されているセンサを用いて高齢者の転倒を検出する手法を提案している。また、大内ら [7] は家庭内におけるセンシングについて「歩行」、「作業」、「安静」という基本動作を推定した上で、主な家庭内作業として挙げられる「歯磨き」、「電気シェーバーによる髭剃り」、「ドライヤーの仕様」、「トイレ・手洗い」、「掃除機掛け」、「皿洗い」、「アイロンがけ」というより細かな行動状態の識別も行っている。以上のように、歩くや走るというような基本行動のみならず、転倒や家庭内行動等、多様な行動の識別課題へ議論が展開されている。特に基本行動データセットについては、HASC コーパスのように多数公開されているという背景もある。

次に、所持位置の推定に関する関連研究を示す。Alanezi ら [2] は、スマートフォンの所持位置が行動推定に影響を及ぼすことを示した上で、加速度・ジャイロセンサの値から Nave Bayes [8], J48 決定木 [9] などを用いて所持位置推定を基本行動推定に応用している。推定する位置には「手」や「ズボン前」、「ズボン後」、「ジャケットポケット」の他にも「ビデオ視聴」、「通話時」の端末位置を対象としている。河内らの研究 [10] ではスマートフォンの加速度センサ値のみを用いて多層パーセプトロン [11] を用いて手法提案、評価を行っている。推定箇所として「ズボン前ポケット」、「ズボン後ポケット」、「胸ポケット」、「ジャケットポケット」、「首から提げる」、以上 5 箇所を対象としていた。我々の先行研究 [12] では加速度センサ、ジャイロセンサの他にも照度センサ、近接センサを用いて J48 決定木の他に Random Forest [13], や多層パーセプトロンを深層化した Deep Neural Network (DNN) を用いた手法を提案し、精度向上を目指した。判定箇所には、ズボン前ポケット、ズボン後ポケット、胸ポケット、内ポケット、ジャケットポケット、鞆、手の計 7 箇所としている。ただし、スマートフォンを手に持つ場合は様々な位置での利用が考えられるが、「歩行時に画面を見ること」と「手に持って下げている」状態をまとめて「手」としている。以上のように、所持

表 1: 使用する CNN モデル構造とレイヤ名称

Layer Type	Name	Shape of Filter	Shape of Output	Output Channels
Conv2D	C1	(1,11)	(3,246)	32
Conv2D	C2	(1,9)	(3,238)	128
Max Pooling2D	—	(1,2)	(3,119)	128
Conv2D	C3	(2,5)	(2,115)	256
Max Pooling2D	—	(1,2)	(2,57)	256
Dropout	—	—	(2,57)	256
Conv2D	C4	(2,3)	(1,55)	312
Conv2D	C5	(1,2)	(1,54)	312
Max Pooling2D	—	(1,2)	(1,27)	312
Conv2D	C6	(1,3)	(1,25)	256
Max Pooling2D	—	(1,2)	(1,12)	256
Conv2D	C7	(1,3)	(1,10)	128
Dropout	—	—	(1,10)	128
Max Pooling2D	—	(1,2)	(1,5)	128
Flatten	—	—	—	640
FC	D1	—	—	8192
FC	D2	—	—	4096
FC	D3	—	—	1024
Dropout	—	—	—	1024
Softmax	Output	—	—	6 or 7

位置推定という課題に対しては、各々高性能な学習手法を提案し、推定精度の高精度化を目指しているが、現時点で提案されている手法は手で特徴量を抽出するものが殆どであるという状態であった。所持位置が高精度に推定できると、各種行動推定精度の向上にも繋がるため、より高精度な所持位置推定の実現が切望されている。しかし、所持位置推定の関連研究は先に挙げたような行動認識とは比較的活発には議論なされていないので、公開データセットがない。そこで、自前で収集したデータセットを実験に用いるが、公開データセットの規模と比較すると大規模とは言えない。DNN や CNN のような深層学習は大量のデータを学習に用いることで高精度を達成する学習手法であり、サンプルが多ければそれだけ高性能を発揮するが、自前で用意したデータセットは深層学習を用いても高精度を実現しにくいという背景もある。

最後に転移学習に関する関連研究を挙げる。Donahue ら [4] は ImageNet 学習した CNN モデルを他の画像分類問題の解決に転用する手法を提案している。さらに、自身の提案手法を実験で様々な画像分類課題に適用することで、転移前に対する推定精度の向上を確認し、CNN の中間層より抽出される特徴の汎用性が非常に高いことを示した。また、Andrej ら [14] のように動画像に移っている人の行動を識別する研究もなされており、彼らも画像を学習したモデルによる転移学習が、動画像内の人の行動認識精度を大幅に高めることを示している。彼らを始めとして、画像分野への転移学習の有効性を示している研究は盛んに議論されている。一方、時系列センサデータの転移に関する研

究は Hayashi ら [15] や Matsui ら [16] のように家庭内基本行動の課題等で試されているが、特に、行動推定を SD、所持位置推定を TD とした研究は私の調べた限りでは見つからなかった。そこで、本研究では行動推定より学習したモデルをスマートフォン端末所持位置推定に転移することの有用性を検証する。

関連研究の手法を踏まえた本研究の位置付けは、所持位置推定を行う CNN モデルを開発すること、並びに、行動認識におけるベンチマークデータセットを用いて同モデルを訓練し、これを転移させることによる推定精度の変化を考察することである。

3. 提案手法

3.1 提案手法の概要

本稿で開発した CNN モデルの概要を表 1 に示す。層の名称について、畳込み層は Conv2D、Flatten は畳み込みによって出力される 2 次元ベクトルを 1 次元へと変換する層である。Max Pooling2D、Dropout 等は名称通りの役割を果たす。FC は全結合層を指す。Conv2D と FC に関しては後の評価実験のため簡易な名称を、浅い方から順に C1、C2 と名付けている。出力層の活性化関数には Softmax 関数を用いている。また、Dropout 層に関しては Dropout Rate は全て 0.5 としている。このモデルは、予備実験として、事前にチューニングした結果、所持位置推定の推定結果について高精度を達成するように探索的に求めたモデルを基にした。次に転移学習を応用する場合の学習手法を示す。本稿で提案する転移学習手法は図 1 に示すように、

表 2: 被験者 12 名の詳細

ID	Age	Gender	Devices	Tops	Bottoms	Shoes	Bags
A	23	男	Nexus 5	スーツ	スーツ	革靴	ブリーフケース
B	23	男	Nexus 5	スーツ	スーツ	革靴	ブリーフケース
C	26	男	Galaxy Note 3	テラードジャケット	ジーンズ	スニーカー	ブリーフケース
D	29	男	Galaxy S2	スーツ	スーツ	革靴	ブリーフケース
E	31	男	Nexus 5x	スーツ	スーツ	革靴	ブリーフケース
F	33	男	Xperia Z3 Compact	テラードジャケット	ジーンズ	スニーカー	リュックサック
G	25	女	Galaxy S2	マウンテンジャケット	ジーンズ	スニーカー	リュックサック
H	27	女	Galaxy S2	トレチコート	ジーンズ	ヒール	ショルダーバッグ
I	27	女	Xperia A4	スーツ	スーツ	靴下	バケットバッグ
J	27	女	Nexus 5X	トレチコート	チノパン	ヒール	バケットバッグ
K	29	女	Galaxy S2	テラードジャケット	スカート	ヒール	ショルダーバッグ
L	36	女	Galaxy S2	白衣	ジーンズ	サンダル	トートバッグ

SD となるタスクを行動認識, TD となるタスクを所持位置推定として転移学習を行う. 両ドメイン共に入力は加速度センサデータ xyz の 3 軸*256 サンプルとし, 出力は各ドメインの分類対象とする. なお, 詳細は後述するが, 評価実験ではこのモデルについて部分的な転移を試すことによって精度向上を図るため, 今回は昨今発表されている ResNet [17] のような特に深層な構造の推定モデルは用いていない.

本研究における行動認識は 6 種類の分類 (Stay (停止), Walk (歩く), Jog (走る), Skip (スキップ), StUp (階段上る), StDown (階段降り)) であり, 所持位置推定は 7 種類の分類 (ズボン (前, 後ろ, ジャケット, 内, 胸), 手, 靴) である. よって, 行動認識モデルの出力層のユニット数は 6 としているが, これを所持位置推定に転移する場合には所持位置推定モデル出力層のユニット数を 7 に変えて, 出力層以外の層の重みを転移させる. ドメイン間で解く問題は異なるものの, 3 軸加速度センサデータを入力としており, 着目すべき特徴が類似している可能性があるため, 学習データの疑似的な拡充ができ, TD の推定精度の向上が期待される. 本研究では転移学習の有無による TD の推定精度の向上を確認すると共に, 転移する層の変化に伴う推定精度の変化を評価実験によって明らかにする.

3.2 使用するデータセット

行動認識モデルの入力に用いるデータセットは HASC コーパス [1] を用いる. 本研究では, 利用する計算機の都合上, HASC コーパスに記録されている者の中よりサンプル数の多い者をピックアップした 19 人の加速度センサデータ約 100 万サンプルを用いる. なお, このデータはサンプリング周波数 100Hz で記録されているデータを抽出して利用している. 256 サンプルを 30 サンプルずつスライドさせて作ったものを 1 フレームとし, 最終的に生成したフレーム 34911 件を学習や評価に用いる. それぞれのフレームには先述した 6 種類のラベルを振り分けており, 行動の

分類ができるデータとなっている. HASC コーパスの所持位置はズボンポケットや腰の位置に所持しているデータを学習に用いている. 端末は iPhone3GS や iPod touch 3G, 小型無線加速度センサ WAA-001 等が用いられている.

所持位置推定モデルの入力には先行研究 [12] で収集したデータセットから, 生データが統一的に取得できた 12 名のデータを使用する. 収集したデータからフレーム 16,153 件生成し, 学習評価に用いた. データ収集の際は, まず加速度センサを記録するアプリケーションを作動させた後, 各所持位置にスマートフォンを格納し歩行してもらう. なお, 端末格納方向としては, スマートフォンを上向きで格納する場合と下向きで格納する場合との 2 種類で複数セッション実施している. 歩行は各 1 分程度行い, その間に U ターンやカーブも含んでいる. 被験者 12 名の詳細を表 2 に示す. 利用機種や服装が異なっており, それぞれ異なる場所で実験を行っている. 研究によっては, 端末や服装のような条件を揃えたデータを用いている場合があるが, 実環境における提案手法の推定精度の評価として適切でないことが懸念される. 本研究においては, 端末や服装などがある程度異なっている状況でどの程度の推定精度が実現できるかを確認したいため, 条件は敢えて揃えていない.

4. 評価実験

4.1 評価手法

提案手法の有効性を確認する為に, 本研究では作成したモデルに対して精度の評価を行う. 精度評価について, 自己データを含めて学習してしまうと推定精度が著しく向上してしまうことが知られている [10]. そこで, 本研究では Leave-One-Out-Subject Cross Validation (LOSO-CV) を用いて精度評価を行う. LOSO-CV では, テストデータを被験者で分けて精度評価を行う手法である. 本研究ではテストデータを 1 人分とし, 残りの 11 人分のデータを学習に用いる. 評価指標としては, クラス間のデータ数も大きく偏っていないことから, Accuracy を用いる.

表 3: 特徴量の名称

Feature Names			
平均値	動きの激しさ	低周波 FFT 標準偏差	中周波 FFT 四分位範囲
絶対値の平均値	周波数パワースペクトル (FFT) の最大値	低周波 FFT 第一四分位数	中周波 FFT 軸間の相関係数
標準偏差	FFT 最大の周波数	低周波 FFT 中央値	高周波 FFT の最大値
絶対値の標準偏差	FFT 最小値	低周波 FFT 第三四分位数	高周波 FFT 最大の周波数
最小値	FFT 標準偏差	低周波 FFT 四分位範囲	高周波 FFT 最小値
最大値	FFT 第一四分位数	低周波 FFT 軸間の相関係数	高周波 FFT 標準偏差
二乗平均平方根	FFT 中央値	中周波 FFT の最大値	高周波 FFT 第一四分位数
第一四分位数	FFT 第三四分位数	中周波 FFT 最大の周波数	高周波 FFT 中央値
中央値	FFT 四分位範囲	中周波 FFT 最小値	高周波 FFT 第三四分位数
第三四分位数	FFT 軸間の相関係数	中周波 FFT 標準偏差	高周波 FFT 四分位範囲
四分位範囲	低周波 FFT の最大値	中周波 FFT 第一四分位数	高周波 FFT 軸間の相関係数
軸間の相関係数	低周波 FFT 最大の周波数	中周波 FFT 中央値	
軸間の絶対値の相関係数	低周波 FFT 最小値	中周波 FFT 第三四分位数	

表 4: 学習手法別精度評価 (太字は手法中最高値)

ID	DNN	転移前 CNN	転移後 CNN (C5)
A	75.4	64.4	79.5
B	65.0	68.0	76.2
C	61.6	56.9	67.8
D	73.8	65.6	67.0
E	47.5	41.0	51.9
F	53.9	39.4	33.2
G	55.0	63.3	55.7
H	78.6	60.4	81.2
I	48.7	36.0	54.2
J	71.4	59.7	60.1
K	45.4	51.5	64.8
L	58.2	48.7	55.3
Total	60.3	54.2	64.9

なお、比較対象として示す DNN については、先行研究 [12] で使用されている特徴量をセンサ生データより抽出して入力に用いている。その特徴量を表 3 に示す。これらの特徴量を各 3 軸に対して抽出し、利用する。低周波数・中周波数・高周波数パワースペクトルについては、加速度センサの波形を基にピーク付近を抽出する為、周波数パワースペクトルの内 0.4~1.6Hz 帯を低周波数パワースペクトル、1.6~3.1Hz 帯を中周波数パワースペクトル、3.1~4.7Hz 帯を高周波数パワースペクトルとし、それぞれから通常の周波数パワースペクトルと同じ特徴量を抽出した。また動きの激しさは標準サイズを n とし、 x 軸センサ値を $x_i (i: 1, 2, 3, \dots, n)$ とするとき、以下の式で表され、動きの激しさ (Intensity) を定義している。

$$Intensity = \frac{1}{n-1} \sum_{i=1}^{n-1} |x_i - x_{i+1}| \quad (1)$$

4.2 推定精度比較

先行研究において現状有効性を示されている DNN と、

CNN 転移前後の精度を確認した。比較結果を表 4 に示す。太字で表した推定精度は、同一被験者内において達成した最高推定精度を示している。また、転移学習した後の推定精度は、後述する実験において最良の結果であった C5 層まで転移させた場合の精度を記載している。また、表最下部の Total は、全推定結果を結合したときの全体の推定精度を示している。表 4 より、CNN を用いても転移学習に行動認識モデルを用いない場合は DNN の推定精度には及ばず、Total でも 6.1 %低い結果となった。よって、CNN のような表現学習手法を単純に用いた状態では、これまでの知見に基づいた特徴量以上に有用な特徴は見出せていないと言える。特に、歩行時の所持位置推定という課題では、同じ歩行動作の中の所持位置の違いによるセンサデータの波形の細かな違いを発見することが必要になっており、それが転移前のデータセットの規模では難しかったのではないかと考察される。しかし、転移学習を用いた場合は精度が向上し、転移前よりも Total で 10%程の精度向上が確認された。被験者の中には 10%以上の精度向上が確認されたものもある。さらに DNN の推定結果よりも 4.6%高い精度を実現できている。従って、転移学習によって学習データを拡充して CNN モデルを形成したことは効果的であったと言える。ただし、被験者 D や F, J, L 等、一部 DNN の方が高い精度を達成する場合もある。これらの被験者には、CNN による特徴抽出が効果的ではなかったと考えられる。手動で特徴抽出する DNN の方が高い精度を出していることもあり、有効な特徴量等の個人による差異が影響していると考えられる。その為、全ての被験者に対して CNN が有効であるとは言えない。それでも、転移学習を用いた CNN を用いると最も高い場合が被験者の半数以上を占めており、Total としても高い精度を達成している。以上より、公開センサデータセットを用いて行動認識を学習してから所持位置推定モデルを学習すると、多くの場合で高精度な推定モデルを生成できると結論付けられる。

表 5: 転移前の CNN モデルによる被験者を集約した推定結果内訳 (front～hand は [件], Recall と Precision は [%])

Cor. \ Pre.	Front	Back	Chest	Inner	Jacket	Bag	Hand	Recall
Front	1587	308	164	79	514	40	9	58.8
Back	422	1552	459	297	178	25	42	52.2
Chest	24	184	1265	343	104	29	15	64.4
Inner	36	162	754	480	190	40	6	28.8
Jacket	178	171	483	560	1114	212	114	39.3
Bag	83	66	105	58	331	888	59	55.8
Hand	9	126	131	3	201	88	1871	77.0
Precision	67.8	60.4	37.6	26.4	42.3	67.2	88.4	54.2

表 6: 転移後の CNN モデルによる被験者を集約した推定結果内訳 (front～hand は [件], Recall と Precision は [%])

Cor. \ Pre.	Front	Back	Chest	Inner	Jacket	Bag	Hand	Recall
Front	1932	278	45	79	333	18	15	71.6
Back	299	2216	131	198	101	13	17	74.5
Chest	49	248	1254	202	185	6	20	63.8
Inner	60	242	632	586	115	32	1	35.1
Jacket	236	298	245	445	1499	68	41	52.9
Bag	30	38	18	208	298	937	55	59.2
Hand	21	71	120	28	104	26	2060	84.8
Precision	73.5	65.3	51.3	33.6	56.9	85.2	93.3	64.9

次に、表 5, 6 に全被験者の結果を集約した CNN による推定結果転移前後の Confusion Matrix を示す。表 5 より、転移前の段階から前ポケットや後ろポケット、鞆や手に所持している場合は加速度センサデータより分類しやすいが、胸ポケットや内ポケット、またジャケットポケットの誤分類が多いことが分かった。中でも表 5 中の太字で示したように内ポケットの推定精度が極端に悪くなっている。これはジャケットの内ポケットと胸ポケットの位置が他の分類対象よりも近い位置にあることが原因で、センサデータにもあまり違いがなかったため、このような結果になっている。しかし表 6 に示すように、転移学習を用いた場合は推定精度の改善が見られる。特に本来 Jacket であるデータが Bag や Hand へと予測されていたケースが減少するなど、全体的に誤分類が減少した。その影響で、特に Jacket のデータは Inner に誤るケースは若干の減少に留まるものの他クラスへ誤るケースが減少し、Precision, Recall 共に 10% 程向上している。その他の部位についても同様に誤分類の減少が見られ、推定精度向上に繋がったと思われる。

転移学習により、精度改善は見られ、先行研究 [12] よりも高い精度を示すことが確認されたが、達成できた精度は最大でも被験者 H の 81.2% であり、未だ実用には至らないと思われる。従って、さらなる改善が必要である。同じ手法を用いて DNN を利用した場合の精度に関して、先行研究の文献内の推定精度より低い値であった。この原因も考慮に入れて改善に努めるべきであると考えている。その原因としては以下が考えられる。

(1) 先行研究でも使用したデータを本研究でも利用したと

は言え、本研究で利用したデータが限定的であった。

(2) 本研究では加速度センサのみ、さらに、生データをそのまま入力に用いている

1 点目の原因については、生データとして統一的に揃えて取得できなかった被験者を本研究では除いていた。この除外した被験者の中で学習に有用なデータがあった可能性がある。2 点目については、本研究では重力成分と加速度成分に分けることなく生データを入力していることや、ジャイロセンサ等他種センサを用いていないためであると考えている。いずれにしても、今後は除外した被験者のデータとの違いや、使用センサの種類による違い、CNN が抽出した特徴や、DNN の方が有効であった場合の有効な特徴量など、データの性質についても検証を深めていき、所持位置推定モデルの精度向上に繋げていきたい。

4.3 転移部分による推定精度変化

転移させる層によってもモデルの性能が変化することが知られている [15, 16]。そこで、本研究では行動認識モデルの重みを共有する層を変えることによる推定精度の変化を調べ、部分的な転移の有効性を確認した。本研究では、図 2 のように、全くモデルを転移させずに新規学習を行ったモデル (Normal) と、層の浅い方から順に層の重みを複数層共有させて学習を行ったモデルとを精度比較する。図 3 にその結果を箱ひげ図として示す。▲は平均値を示している。図 3 より、浅い層から C4, C5 までを転移させる場合が多く数の被験者にとって最良の結果となった。ただし、特に C4 や C5 まで転移させた場合、精度分布の下限も低く

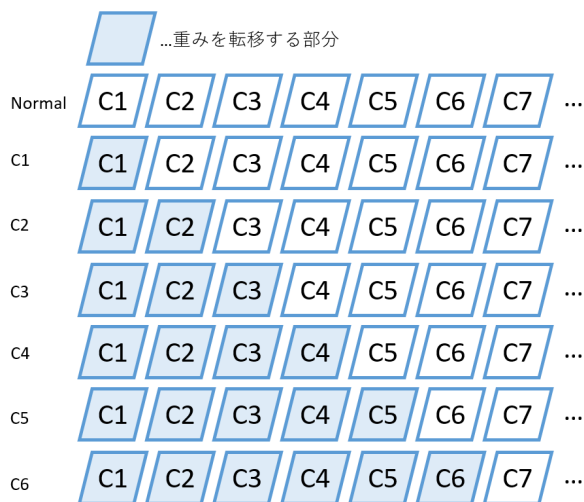


図 2: 行動認識モデルと重みを共有する場所

なっている通り、逆に精度が下がっている場合もあるため、転移させる最適な層にも個人差が存在することが確認できる。それでも中央値としては C4 や C5 が他の全てより高い水準であり、平均値としても高い精度分布にまとまっているのは C5 である。そこで、今回作成したモデルの場合は浅い C1 から C5 まで転移させる手法が最適であると結論付け、これを表 4 の転移後の結果とした。この結果のように、全てを転移させるよりも約半分程転移させる手法の方が最高精度を達成することもあるため、転移学習を利用する場合はこのような視点でのチューニングも必要となる。

5. おわりに

本研究では、所持位置推定モデル作成に CNN を応用し評価した。加えて、大量のデータで学習した行動認識モデルの重みを、所持位置推定モデルに転移学習させる手法の有効性も確認した。CNN は主に画像処理の領域で高い性能を発揮しているが、入力センサデータとなる場合にもある程度有効に働く。本研究の所持位置推定モデルの評価では、先行手法よりも Total で約 10% の精度向上が見られ、被験者の中には 10% 以上向上が確認された者もいる。

また、本研究ではモデル全体的な転移の他に、部分的に転移する場合の有効性も確認した。結果としては、本研究で用いたモデルでは、モデルの中間辺りまで転移させると最も有効に働き、最高精度を達成した。転移する場所もパラメータチューニングの範囲として考慮すると、精度向上が見られることが分かった。

転移学習の有効性と、提案手法が先行手法よりも高精度となる場合があることが確認されたが、提案手法の推定精度としては最大でも 81.2 % であり、実用の為にはさらなる精度向上が必要である。その為には、胸ポケットや内ポケットを同一扱いにするといった分類粒度の再検討や、先行研究で利用し、本研究で使用しなかったデータの量、使用するセンサの種類等、データに着目した改善手法の検証

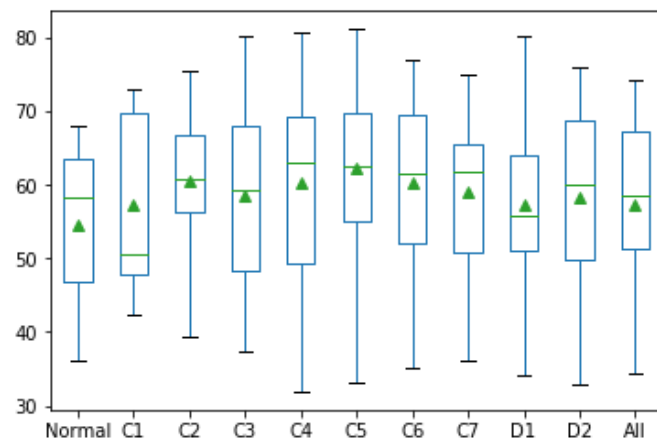


図 3: 転移部位別の精度分布

が必要である。所持位置推定のような、分類が難しい問題に対しては個人差の影響を考慮したユーザのクラスタリングや、クラスタリング結果を利用したアンサンブル学習の導入も検討していくべきであると思われる。今後はこのような改善手法を試し、精度向上を目指していく。

謝辞

本研究の一部はスズキ財団の科学技術研究助成によるものである。ここに謝意を示す。

参考文献

- [1] Nobuo Kawaguchi, Nobuhiro Ogawa, Yohei Iwasaki, Katsuhiko Kaji, Tsutomu Terada, Kazuya Murao, Sozo Inoue, Yoshihiro Kawahara, Yasuyuki Sumi, and Nobuhiko Nishio. Hasc challenge: gathering large scale human activity corpus for the real-world activity understandings. In *Proceedings of the 2nd augmented human international conference*, p. 27. ACM, 2011.
- [2] Khaled Alanezi and Shivakant Mishra. Design, implementation and evaluation of a smartphone position discovery service for accurate context sensing. *Comput. Electr. Eng.*, Vol. 44, No. C, pp. 307–323, May 2015.
- [3] Jia Deng, Wei Dong, Richard Socher, Li jia Li, Kai Li, and Li Fei-fei. Imagenet: A large-scale hierarchical image database. In *In CVPR*, 2009.
- [4] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In *International conference on machine learning*, pp. 647–655, 2014.
- [5] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L. Reyes-Ortiz. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In José Bravo, Ramón Hervás, and Marcela Rodríguez, editors, *Ambient Assisted Living and Home Care*, pp. 216–223, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.
- [6] Stefano Abbate, Marco Avvenuti, Francesco Bonatesta, Guglielmo Cola, Paolo Corsini, and Alessio Vecchio. A smartphone-based fall detection system. *Pervasive and Mobile Computing*, Vol. 8, No. 6, pp. 883–899, 2012.
- [7] 大内一成, 土井美和子ほか. 携帯電話搭載センサによるリ

アルタイム生活行動認識システム. 情報処理学会論文誌, Vol. 53, No. 7, pp. 1675–1686, 2012.

- [8] Andrew McCallum, Kamal Nigam, et al. A comparison of event models for naive bayes text classification. In *AAAI-98 workshop on learning for text categorization*, Vol. 752, pp. 41–48. Citeseer, 1998.
- [9] J Ross Quinlan. *C4. 5: programs for machine learning*. Elsevier, 2014.
- [10] 河内智志, 薛媛, 藤波香織. 携帯端末の身体上格納場所判定機能のスマートフォンへの実装. *インタラクシオン*, Vol. 2011, pp. 531–534, 2011.
- [11] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, Vol. 65, No. 6, p. 386, 1958.
- [12] 長谷川達人, 越野亮ほか. 深層学習を用いた歩行時におけるスマートフォンの所持位置推定. 情報処理学会論文誌, Vol. 57, No. 10, pp. 2186–2196, 2016.
- [13] Leo Breiman. Random forests. *Machine learning*, Vol. 45, No. 1, pp. 5–32, 2001.
- [14] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 1725–1732, 2014.
- [15] Tomoki Hayashi, Masafumi Nishida, Norihide Kitaoka, Tomoki Toda, and Kazuya Takeda. Daily activity recognition with large-scaled real-life recording datasets based on deep neural network using multi-modal signals. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. 101, No. 1, pp. 199–210, 2018.
- [16] Shinya Matsui, Nakamasa Inoue, Yuko Akagi, Goshu Nagino, and Koichi Shinoda. User adaptation of convolutional neural network for human activity recognition. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pp. 753–757. IEEE, 2017.
- [17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.