

行動認識モデルの転移学習に向けた ニューラルネットワークによる特徴抽出の可視化と分析

吉村 直也¹ 前川 卓也¹ 原 隆浩¹

概要：

行動認識技術はコンテキスト依存アプリケーションのための基礎技術である。深層学習はデータからの特徴学習を可能とし、その認識精度も優れているため行動認識分野においても注目を集めている。その中でも転移学習は、ラベル付きデータの量が少ない場合でも高い精度で認識することができる可能性をもつ重要な技術の一つであり、近年盛んに研究されている。しかし行動認識データセットはセンサの種類や装着位置・認識対象の行動クラスが異なる事が多いなど、想定する利用環境と同じ想定で取られたデータセットを探す事は難しい。このような問題を解決すべく研究が行われているが、包括的に解決する手段はまだ提案されていない。本研究ではこの問題を解決し、転移学習の行動認識分野への応用を促すため、行動認識モデルの分析と転移学習を行う方針に関して考察を行う。行動認識モデルの分析を行うため、ニューラルネットワーク中のユニットが最も大きな出力を出す入力を用いた勾配上昇法を用いて計算する。ハイパスフィルタを用いた正則化を提案し、高周波数成分が少ない行動認識モデルに適した信号を生成する。また、モデルの出力である softmax 関数の出力結果に基づいて可視化を行なう。この可視化により各ユニットがどの行動クラスの認識に寄与しているか、また転移学習にどの重みが利用できるかを考察する。この可視化の結果に基づいて、認識精度が高くなる転移学習の方法を考察する。実際に転移学習を行なってその有効性を確認した。

1. 概要

近年、健康志向の高まりに伴いウェアラブルデバイスが普及している。ユーザは日常的にこれらのデバイスを装着し、日々の行動やエクササイズの内容を記録することで健康増進や運動の質の向上に活用している。行動認識技術はウェアラブルデバイスから取得したセンサのデータから行動の内容を推定するための技術であり、コンテキストウェアアプリケーションの基礎的技術として盛んに研究されている。

深層学習は、データからの特徴学習を可能とし、その認識性能も優れているため、行動認識の分野においても注目を集めている。深層学習の画像分野における目覚ましい発展の背景には、大規模データセットと VGG などの事前学習したモデルを用いた転移学習があると考えられる。前者は汎化性能が高い特徴抽出を可能とし、後者は学習時間の削減と小規模データセットへの応用に大きな役割を果たしている。

しかしながら行動認識の分野における転移学習の利用は容易ではなく、下記のような問題が存在する。

- データセットによってセンサの装着位置が異なるなど、データ収集環境の差異によって想定する環境で転移学習を行えるデータセットを見つけることが難しい。
- 認識対象の行動クラスもデータセットで異なり、深い層の学習済みパラメータを転移することが難しいため、転移学習を行うメリットが低い。

既存の行動認識に関する研究においては、想定する環境で研究者自身が大量のデータ収集をする必要があり、多大な手間がかかる場合が多い。行動認識分野において転移学習を有効に活用できれば、少量のデータで提案手法の検証が可能になり、研究のスピードを加速させることができる。センサの装着位置の違いなど個々の問題に対する取り組みは近年行われているが、異なるデータ収集環境での転移学習を行うためには、これらの問題を包括的に解決する手法が必要である。

そこで本研究では、行動認識分野における転移学習の有効利用に向けた、認識モデルの分析と転移学習を行う方針に関する考察を行う。そのため、まず行動認識のニューラルネットワークがどのような特徴量を学習しているのかを可視化し、異なるデータセット間で学習されている特徴の差異を検証する。この可視化とドメイン間の比較を行うこ

¹ 大阪大学大学院情報科学研究科

とで、センサの装着位置や対象とする行動クラスにおいて共通する特徴の存在や、データセットに特化した特徴の存在に関して知見を得ることができる。これは転移学習や、先に述べた問題を解決する包括的な手法を考案する上で重要な判断材料である。さらにその検証結果に基づき、ドメイン間とセンサの装着位置間での転移学習方法の指針について考察する。実際に指針に基づいて転移学習を行い、その有効性を確認する。

本研究ではCNNを用いたニューラルネットワークを対象として可視化を行う。学習された特徴は、(1) ネットワーク中の各ユニットが最も強く反応する信号と、(2) ユニットの出力の強さと行動ラベルの関係性の2つに着目して可視化を行う。後述する勾配上昇法を用いて注目するユニットが最も大きな出力値を出す入力信号を求め、特定のユニットが学習した内容の可視化を試みる。勾配上昇法を用いると高周波成分が非常に強く現れてしまう問題があるが、人間の行動は比較的低周波数帯で行われる。そこで本研究では高周波成分を抑制する正則化手法を提案することで人間の動作との関連性が低い高周波成分を抑制し、より行動認識モデルの理解により有効な入力信号を生成を試みた。また各ユニットの出力が大きい場合に、入力したサンプルがどの行動クラスのサンプルであるのかを調べ、各ユニットと行動ラベルの関係性を可視化を試みた。

本研究の技術的貢献点は以下の通りである。

- 行動認識モデルに対して勾配上昇法を用いて、ネットワーク中の特定のユニットが抽出している学習内容の可視化を行った。行動認識のデータに適した可視化を行うため、ハイパスフィルタを用いた正則化項を提案する。
- 行動認識モデルの可視化手法を適用し、データセット・センサの装着位置の違いにより学習される特徴の傾向を比較した。またこの結果に基づいた転移学習の適用方針を考察し、実際に転移学習を行って検証した。

まず第2節において、ネットワークの可視化手法および行動認識の転移学習に関する研究を紹介する。第3節で本研究で用いる可視化手法の詳細を説明し、これを用いて第4節でデータセット間で比較を行い、学習データに起因する学習傾向の相違に関して考察を行う。また考察結果に基づいて転移学習の方針に関して考察する。第5節において実際に転移学習を行い、第4節における考察の有効性を確認する。

1.1 想定する転移学習

本研究では以下のような条件に該当する転移学習を想定する。

- 転移元・転移先環境に関しては、共にラベル付きのデータが存在する。ただし転移先環境で少量のデータのみラベルが付与されている場合を想定する。

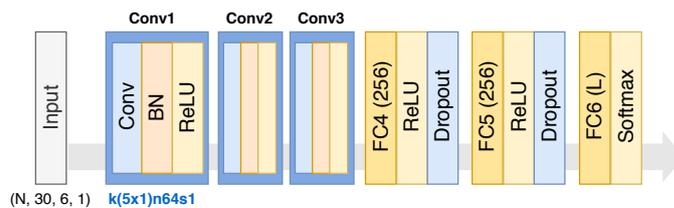


図1 行動認識ネットワーク (Lは認識する行動クラス数)

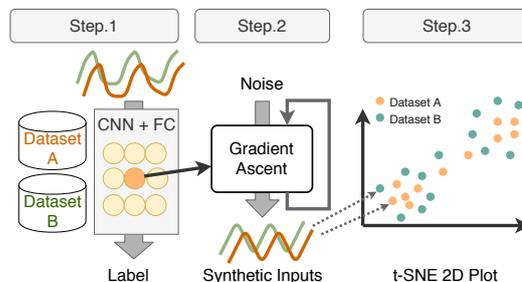


図2 CNNフィルタによる抽出特徴の可視化手法の概要

- (1) ドメイン (データセットおよび行動クラス) 間および、(2) センサの装着位置間の転移学習を行うことを想定する。
- 身体部位に装着された加速度センサを使用する。

2. 関連研究

2.1 ニューラルネットワークの可視化

ニューラルネットワークの内部は基本的にはブラックボックスとみなされるが、このネットワークの内部処理を理解するために様々な方法が提案されている。まずニューラルネットワークの層の粒度で、学習結果の理解を試みる研究を紹介する。Yosinskiら[1]は、注目している層が一般的な特徴と転移元環境に特化した特徴のどちらを抽出しているかを可視化を行なった。まず学習済みの重みを層ごとに別のタスクに転移学習を行う。この結果得られる転移先タスクにおける精度を基に、注目する層が一般的な特徴もしくは、学習データに特化した特徴を抽出しているかを定量化した。また転移学習を用いずに、ニューラルネットワークの出力値を入力信号に逆変換することで、ネットワーク中のそれぞれの層の処理内容を可視化する研究も行われている。Mahendranら[2]は、注目する層の出力から入力画像を推定するニューラルネットワークを学習することで、レイヤーレベルで出力値を入力画像へと逆変換する手法を提案した。

ニューラルネットワーク中の層ではなく、各ユニットに注目して学習内容を可視化する研究も行われている。Yosinski[3]らは勾配上昇法を用いて、あるノードの出力を最大化する入力信号の生成に取り組んだ。またより解釈性が高い入力信号を生成するために、L2正則化や認識にあまり寄与しない画素をクリッピングするなど、様々な正則

表 1 行動クラス

Opportunity (LOC/GES)		PAMAP2 (PAM)		CSLab (CSL)	
L1: Stand	G8: Toggle_Switch	1: Lying	12: Vacuum cleaning	1: Standing	12: WritingNotebook
L2: Walk	G9: Open_Dishwasher	2: Sitting	13: Ironing	2: Walking	13: PlayPingpong
L3: Sit	G10: Open_Drawer3	3: Standing walking	14: Folding laundry	3: Running	14: DrawingWhiteboard
L4: Lie	G11: Open_Drawer2	4: Running	15: House cleaning	4: Sitting	
G1: Close_Dishwasher	G12: Open_Door1	5: Cycling	16: Playing soccer	5: StairsUp	
G2: Close_Drawer3	G13: Open_Door2	6: Nordic walking	17: Rope jumping	6: StairsDown	
G3: Close_Drawer2	G14: Open_Drawer1	7: Watching TV		7: Bicycling	
G4: Close_Door1	G15: Open_Fridge	8: Computer work		8: BrushingTeeth	
G5: Close_Door2	G16: Drink_Cup	9: Car driving		9: WashingDishes	
G6: Close_Drawer1	G17: Clean_Table	10: Ascending stairs		10: UsingPC	
G7: Close_Fridge		11: Descending stairs		11: Vacuuming	

化方法を提案している。本研究ではより細かい粒度で行動認識モデルを理解するために層ではなく各ユニットに注目し、勾配上昇法を用いて可視化を行う。Yosinski らは勾配上昇法を用いると高周波成分が強調されて入力信号に現れることを指摘している。人間の動作は比較的low周波な動作が多く、本研究が対象とする行動認識で高周波成分が強調されると解釈の妨げとなる。提案手法では、周波数フィルタを用いた正則化手法を提案する。

2.2 転移学習のポリシー

転移学習はあるドメインで学習させたモデルを別のドメインに適応させる手法であり、転移学習は顔認識や画像の高解像度化、ロボットの音声制御などの画像や自然言語処理など多くの分野で成功をあげている [4], [5], [6]. Imagenet など膨大なラベル付きデータで学習した AlexNet[7] などの CNN を用いたモデルは、入力層に近い層では汎用的な特徴が学習されていることが経験的に知られている。この特徴は他のタスクにおいても有効であり、この学習済みの重みを転移することで学習時間の削減や、転移先環境のラベル付きデータが少ない場合に簡単に高い精度のモデルを作ることができる。

転移学習を行うためには、転移する学習済みパラメータの選択と利用方法を決定する必要がある。学習済みパラメータの利用方法とは、(1) 転移したパラメータを固定して転移先モデルでそのまま使用する方法と、(2) 転移したパラメータを含めて低い学習係数を用いて再度学習する方法がある。しかしこれらの学習済みパラメータの利用方法は、経験や試行錯誤に基づいて決定されることが多い、コンピュータビジョンの分野においてはこの経験的な知見が多く共有されているが、行動認識分野では転移学習がまだ広く用いられていないためこのような経験的な知見が乏しい。本研究では学習内容の可視化を通して、行動認識モデルにおいて転移学習を行う方針に関して検討する。

2.3 行動認識におけるニューラルネットワークの利用と転移学習

ニューラルネットワークは、データからの特徴学習を可能とし、その認識性能も優れているため、行動認識の分野においても盛んに研究が行われている。Francisco ら [8] は、マルチモーダルな行動認識の実現に向けて CNN と LSTM を

組み合わせた Deep Convolutional Recurrent Neural Network を提案した。また Münzner [9] らは、マルチモーダルな環境において、ネットワークの出力層の近くでセンサフュージョンを行う Late Fusion の有効性を報告した。本研究では、CNN をベースとしたニューラルネットワークを用いて行動認識を行う。また転移学習を行うにあたって、センサへの依存度が低い特徴を学習することが望ましいと考えられる。本研究ではセンサフュージョンの方法として、Münzner らが提案した全てのセンサで畳み込みフィルタを共有する Shared-Filter Hybrid Fusion を採用する。

2.4 行動認識データセット

行動認識アルゴリズムの開発・評価のためのベンチマークとして利用されるデータセットが多く存在する。Roggen ら [10] が作成した「OPPORTUNITY Activity Recognition Data Set」は、一般家庭のキッチンを再現した環境において 4 名のユーザの行動を記録したデータセットである。ユーザには 7 個の IMU と 12 個の加速度センサが装着され、行動・コンテキスト認識、セグメンテーション、センサフュージョンのベンチマークとして利用できる。また Reiss ら [11] が作成した「PAMAP2 dataset」は、家の中での行動からエクササイズなど 18 種類の多岐に渡る行動を記録したデータセットであり、行動認識タスクのベンチマークとして活用されている。前川ら [12] は、教師なし行動認識技術の評価のために、約 60 名の被験者からより自然な状態での行動データを記録し、大規模なデータセットを構築した。本研究においてこのデータセットを「CSLab dataset」と呼ぶ。また、一般的な行動だけでなく病気の症状を検出するために作成されたデータセットもある。Bachlin ら [13] は、パーキンソン病患者の麻痺症状の検出を行う「Dapnet Gait dataset」を作成・公開している。

このように行動認識データセットは数多く作成・公開されているが、後に示すようにそれぞれのデータセットで行動クラスやサンプル数・被験者数が大きく異なる。本研究ではこれらのデータセットを用いて、転移学習に利用しやすい行動クラスやサンプル数などに関して分析・検討を行う。本研究で用いるデータセットとその行動クラスを表 1 にまとめた。

3. 行動認識ネットワークによる抽出特徴の可視化

3.1 行動認識モデル

本研究で使用する行動認識モデルを図1に示す。Moresら[14]とMünznerら[9]のモデルを参考にし、畳み込み層3層と全結合層3層の合計6層によって構成する。またより転移学習に適したモデルとするため、畳み込みフィルタのサイズを $f \times 1$ とすることで全てのセンサ・チャンネルにおいて畳み込みフィルタを共有する。 f は時間方向の長さである。本研究では $f = 5$ を用いた。

3.2 CNN フィルタによる抽出特徴の可視化

3.2.1 可視化手法の概要

転移学習を行うにあたり、ネットワーク中のユニットの処理内容を把握することは、転移すべき層またはユニットを決定する上で重要な判断基準となりうる。例えば転移元のデータで学習したユニットが抽出する特徴が非常に偏っていた場合、転移学習が成功する可能性が低いと判断することができる。

抽出特徴の可視化手法の概要を図2に示す。ユニットが学習した内容の可視化は、次に説明する勾配上昇法を用いて注目ユニットの出力値を最大化する入力信号を求めると行う。本研究では、入力信号として $\mathbf{x} \in \mathcal{R}^{T \times C}$ の多次元時系列の信号を想定する。 T はタイムステップ、 C はチャンネル数を表す。本研究においては $T = 30$ 、 $C = 6$ を用いる。加速度信号は3軸からなるが、本研究ではそれぞれの軸を1つのチャンネルとみなす。勾配上昇法では式1を満たすような、入力信号を勾配法を用いて求める。

$$\mathbf{x}^u = \arg \max_{\mathbf{x}} (a^u(\mathbf{x}) - \gamma R(\mathbf{x})) \quad (1)$$

$a^u(\mathbf{x})$ は、 \mathbf{x} をモデルに入力した場合の、 u 番目のユニットの出力値を表し、 $R(\mathbf{x})$ は正則化項、 γ は正則化項の重みを調整するハイパーパラメータである。

入力信号を求めるとき、勾配上昇法を用いて、式2を繰り返して適用することで \mathbf{x} を更新していく。

$$\mathbf{x}_t \leftarrow \mathbf{x}_{t-1} + \eta \frac{\partial a^u}{\partial \mathbf{x}} \quad (2)$$

t は更新のステップ数を表す。正則化項は、ユニットの出力値 a_u^{raw} から正則化項を引いた $a_u = a_u^{raw} - \gamma R(\mathbf{x})$ という形で、勾配を計算する前にユニット u の出力から値を減算することで適用する。

ただし、単純な勾配法では \mathbf{x} が収束せずに振動する可能性があるため、更新回数ごとに学習係数を変更するRMSpropを基に、式4を用いて更新を行う。

$$h_t = \alpha h_{t-1} + (1 - \alpha) \frac{\partial a^u}{\partial \mathbf{x}} \quad (3)$$

$$\mathbf{x}_t \leftarrow \mathbf{x}_{t-1} + \frac{\eta}{\sqrt{h_t + \epsilon}} \frac{\partial a^u}{\partial \mathbf{x}} \quad (4)$$

h_t は t 回目の勾配更新における勾配の2乗の移動平均、 α は過去の影響を考慮する量を調整するハイパーパラメータ、 ϵ は分母が0になることを避けるための微小定数、 η は学習係数の初期値である。 $\frac{\eta}{\sqrt{h_t + \epsilon}}$ は t 回目の更新における学習係数となる。本研究では $\eta = 1.0$ 、 $\epsilon = 10^{-7}$ を用いた。

Yoshinskiらの報告によると、勾配上昇法を用いて \mathbf{x}^u を生成すると高周波成分が強く出現する。これを抑制するために正則化項 $R(\mathbf{x})$ が重要な役割を果たす。人間の動作周波数はそれほど高くなく、行動認識には30 Hz程度あれば十分だと言われている[15], [16]。そのため勾配上昇法を行動認識モデルに適用するにあたり、高周波成分を抑制する正則化項の設計は非常に重要である。本研究では以下の4種類の正則化項などに関して、行動認識モデルとの親和性について検討した。

3.2.2 l_p ノルムを用いた正則化手法

l_p ノルムを用いた正則化項は式5で表される。

$$R_{l_p}(\mathbf{x}) = \|\mathbf{x}\|_p = \sum_i |x_i|^p \quad (5)$$

l_p 正則化はニューラルネットワークを学習する上で最もよく利用される正則化手法の1つであり、異常値などを抑制することができる。そのため提案手法のベースラインとして用いる。また提案手法では $p = 2$ とする(L_2 正則化)。

PAMAP2 Datasetを学習したモデルの3層目7番目のユニットに関して、正則化を用いずに生成した \mathbf{x} を図3[a]に、 L_2 正則化を用いて生成した \mathbf{x}^u ($u = 7$)を図3[b]に示す。正則化が何も適用されていない場合、高周波成分大きい信号が生成され、信号の値が激しく動いていることがわかる。入力信号の振幅は1G以下であるのに対し、正則化を何も適用しない場合最大20G程度の値が生成されている。注目するユニットの強く反応する信号が誇張されていたとしても、実際に観測されるデータから大きくかけ離れている。一方で L_2 正則化を適用すると、値がおおよそ2G程度の範囲に抑えられており、実際の信号に近い形となった。しかし2層目3層目と深い層になると高周波成分が増加し、異なるユニットの差分を把握することが難しくなる。上記の結果から特徴的な要素を残しながらも、外れ値を抑えることが必要である。

3.2.3 Total Variation を用いた正則化手法

Total Variation[17]は画像処理において平滑化のために利用されており、時系列信号の場合は式6で表されるように、隣り合う時刻の値の差が大きい時に大きな値をとる。従ってTotal Variationを正則化に用いることで、各チャンネルごとに高周波成分が抑制され滑らかな信号が生成されることが期待できる[2]。

$$R_{TV}(\mathbf{x}) = \sum_{t=0}^{T-1} |x_{t+1} - x_t| \quad (6)$$

Total Variationを正則化に用いて生成した \mathbf{x} を図3[c]

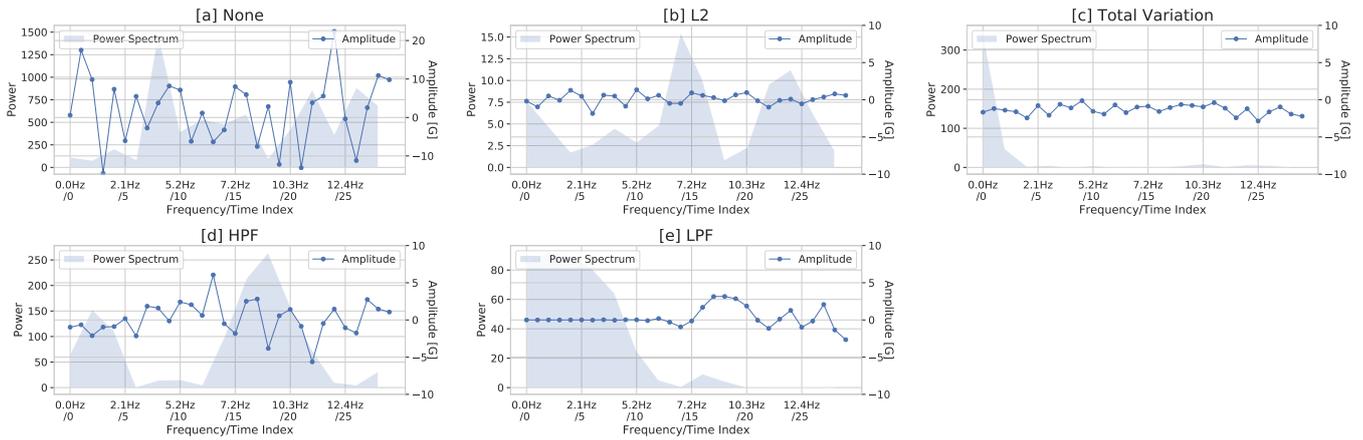


図 3 各正則化手法を用いて生成した \mathbf{x}^u の比較. PAMAP2 dataset を学習したモデルの CNN 3 層目 7 番目のユニットに関して生成した \mathbf{x}^u に関して, 腕の x 軸に当たる信号を実線として, そのパワースペクトラムを塗りつぶし領域として可視化した. また全ての手法において $\gamma = 1.0$ とした.

表 2 学習結果: 各クラスごとに計算したスコアをクラスのサンプル数に基づいて重み付き平均をとったスコア.

評価指標	LOC	GES	PAM	CSL
Precision	0.868	0.441	0.754	0.829
Recall	0.868	0.428	0.746	0.828
F1-score	0.868	0.427	0.743	0.826
サンプル数	6,288	1,605	8,115	111,137

表 3 学習結果: (センサ装着位置)

評価指標	RH-LS (LOC)	LH-RS	LH-LS	RLH	RLS
Precision	0.868	0.864	0.864	0.839	0.870
Recall	0.868	0.863	0.863	0.828	0.869
F1-score	0.868	0.863	0.863	0.826	0.869
サンプル数	6288	6288	6288	6288	6288

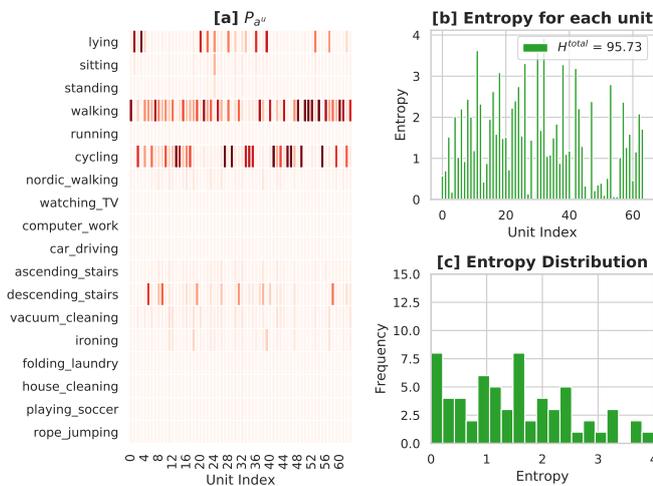


図 4 PAMAP2 dataset を学習したニューラルネットワークの CNN3 層目におけるユニットと行動ラベルの関係の可視化. [a] は全てのユニットに対する $P\alpha^u$, [b] と [c] はそれぞれ各ユニットの対して計算したエントロピーとその分布

に示す. 高周波成分と 2G を超えるような大きな値を抑制できているが, 2 Hz 以下の周波数成分しか残っていない. この傾向は正則化項の強さを調整するハイパーパラメータを変更させた場合も 3 層目のユニットに関しては共通して観測することができる. また 1・2 層目に関して可視化した結果は, 2 Hz 以下と 12 Hz 以上という 2 つの周波数成分が強く出る傾向があった. これは隣接するサンプルの値

の差分を小さくするという正則化をかけた結果, 細かく振動する信号に収束したと考えられる. 今回使用したデータのサンプリングレートでは, 時間方向の滑らかさが十分ではない. したがって, Total Variation のような隣接する値に対してペナルティをかける手法は行動認識モデルに, 適していないと言える. ただし, サンプリングレートが十分に高いデータセットがあれば, 高周波成分の抑制に働く可能性がある.

3.2.4 HPF を用いた正則化手法

より直接的に意図した高周波成分を抑制するために, 周波数フィルタを用いた手法を 2 つ提案する. まず 1 つはハイパスフィルタ (HPF) を用いる方法である. 式 8 のように, 現在の信号 \mathbf{x} に対してハイパスフィルタを適用することで高周波成分のみを取り出し, これに対して l_p ノルムを計算することで勾配に対しペナルティを課す. HPF を用いることで明示的に特定の周波数を抑制することができる.

$$\mathbf{x}^{hpf} = \text{HPF}(\mathbf{x} \cdot \mathbf{w}_{han}, f_e, \delta) \quad (7)$$

$$R_{HPF}(\mathbf{x}) = \sum_i |\mathbf{x}_i^{hpf}|^p \quad (8)$$

HPF(\cdot) はハイパスフィルタ, f_e はカットオフ周波数, δ は遷移帯域幅, \mathbf{w}_{han} はハニングウィンドウを表す. 従ってハイパスフィルタを適用することで, \mathbf{x} から f_e Hz 以上周波数成分のみを取り出すことができる. 提案手法ではパワースペクトルの和ではなく, 時間領域での l_p ノルムを

用いる。提案手法では、時間領域において FIR フィルタを $y_i = \sum_{x=0}^{N_b} b_x x_i$ のように定め、Z 変換を用いて係数 \mathbf{b} を求める。 N_b は \mathbf{b} の長さである。実際のハイパスフィルタの処理はこの係数 \mathbf{b} と \mathbf{x} の積をとることで実現した。提案手法では $f_e = 12.0, \text{delta} = 3.0, p = 1$ を使用した。また今回使用したデータは 30 Hz・ウィンドウ幅 1 sec の信号であり、係数 \mathbf{b} の長さ N_b とウィンドウ内のサンプル数が一致している。より長いウィンドウサイズのサンプルであれば問題ないが、本研究の設定において通常の方法でハイパスフィルタを適用すると、ウィンドウ内のほとんどの値が 0 となる。提案手法では正確な高周波成分のみを抜き出す必要がないので、 \mathbf{x} を時間方向の前後にそれぞれサイズ N_b の 0 パディングを行い、これに対して L1 ノルムの計算を行なった。

ハイパスフィルタを正則化に用いて生成した \mathbf{x} を図 3[d] に示す。高周波成分が抑制され、特徴的な波形がはっきりと現れていることがわかる。また、人間の部位の動きに近い 10Hz 程度の成分が現れていることも分かる。 \mathbf{x}^u の各成分がとる値に 2 G を超えるものが含まれるが、注目しているユニット u の出力を最大化する入力信号が誇張して現れていると考える。全く正則化をかけない図 3[a] などと比較して、ハイパスフィルタを用いた正則化の結果は適当だと考えられる。より高周波成分を抑制したい場合はカットオフ周波数 f_e を調整すれば良いため、様々なデータセットにおいて簡単に意図した結果を得ることができると考えられる。

3.2.5 LPF を用いた信号の修正

周波数フィルタを用いた高周波成分の抑制方法として、ローパスフィルタを用いる方法を提案する。これはハイパスフィルタのように正則化ではなく、毎回の式 4 更新後の \mathbf{x} にローパスフィルタを適用することで、直接 \mathbf{x} から高周波成分を除去する。

ローパスフィルタを正則化に用いて生成した \mathbf{x}^u を図 3[e] に示す。高周波成分が抑制され、滑らかな変化をする \mathbf{x} が生成されている。しかし \mathbf{x} の前半が全て 0 になっている。これはローパスフィルタが x_t の値を計算するとき、 x_{t-N_b}, \dots, x_t までの値を用いるが、 $t < N_b$ の場合その値がなく、計算ができないためである。通常は周波数フィルタの入力の \mathbf{x} が十分長い問題とならないが、本研究のようなデータポイントが少ない場合は無視ができない。ウィンドウ内のデータポイント数が十分に多い場合は、高周波成分を抑制する有効な手段となる可能性がある。

3.3 ユニットと行動ラベルの関係の可視化

3.3.1 可視化手法の概要

3.2 節において、各ユニットが強く反応する信号の可視化を行い、特徴された抽出に関して考察を行えることが示さ

れた。しかし学習した特徴と行動クラスの関係性はわかっていない。本節では各サンプルを入力した時の各ユニットの出力値に基づいて、ユニットとラベルの関係性を可視化する。本研究では、各サンプルのニューラルネットワークによる推定結果の分布に着目し、注目したユニットの各行動クラスに対する相対的な出力値の大きさを可視化する。

サンプル \mathbf{x}_i を入力すると、任意のユニット u から出力値 a_i^u と、出力層から各クラスの推定確率 $\hat{\mathbf{y}}_i = [\hat{y}_{(i,l_1)}, \hat{y}_{(i,l_2)}, \dots]$ が得られる。 l_n は行動クラスを表す。この推定結果を基に、注目したユニットの出力値と行動クラスの関係性を式 10 によって計算する。

$$\hat{\mathbf{a}}_i^u = a_i^u \hat{\mathbf{y}}_i = [a_i^u \hat{y}_{(i,l_1)}, a_i^u \hat{y}_{(i,l_2)}, \dots] \quad (9)$$

$$\mathbf{P}_{a^u} = \text{softmax} \left(\sum_i \hat{\mathbf{a}}_i^u \right) \quad (10)$$

$\hat{\mathbf{a}}_i^u$ は、サンプル i を入力した時のユニット u の出力値を、推定結果を元に各行動クラスに対し分割したベクトルとなる。正解ラベルではなく推定結果を用いて曖昧さを導入することで、行動の境目のサンプルなどにも考慮して注目ユニット u の出力と行動クラスを結びつけることができる。 $\hat{\mathbf{a}}_i^u$ を全サンプルに関して足し合わせ softmax を取ったベクトルであり、要素を全て足し合わせると 1 になる。この $\hat{\mathbf{a}}_i^u$ から行動クラスに対して相対的にどの程度大きい値を出力するかを知ることができる。また \mathbf{P}_{a^u} は、ユニット u のサンプル集合 D_s に対する出力値の、指定した行動クラスへの割り当て量と見ることもできる。

\mathbf{P}_{a^u} はサンプルの内訳に大きく結果が左右される。本研究で使用するデータセットは各行動クラスのサンプル数に大きな偏りがある。したがって本研究では全てのサンプルではなく、アンダーサンプリングを行い全ての行動クラスのサンプルを最も少ない行動クラスのサンプル数に合わせた。可視化結果を図 4[a] に示す。最も左のユニット 0 は「Walk」に強く反応し、隣のユニット 1 は「Lie」に強く反応していることがわかる。

\mathbf{P}_{a^u} の分布が偏っている場合は特定の行動に対して特化された特徴が抽出されており、逆に偏りが弱い場合は汎用的な特徴を抽出していると考えられる。転移学習を考える上でこの抽出される特徴の傾向を知ることが、転移させる重みを決定する上で重要である。そこで偏りを定量化するために \mathbf{P}_{a^u} に対して式 11 に示すエントロピーを計算する。

$$H^u = - \sum \mathbf{P}_{a^u} \log(\mathbf{P}_{a^u}) \quad (11)$$

また本研究では転移学習を層の単位で行うので、注目する層のユニットのエントロピーの総和 $H^{total} = \sum_u H^u$ を定義する。これら可視化結果を図 4[b][c] に示す。ユニット 0-1 は特定の行動に強く反応するため、エントロピーは小さくなっている。また [c] からこの層は幅広いエントロピー

のユニットを持つことがわかるので、特定の行動に特化したユニットと一般的な行動に反応するユニットの両方が存在していると考えられる。

4. 行動認識データセットを用いた分析

前節で紹介した手法を用いて、既存の行動認識データセットで学習したモデルの抽出特徴の可視化と、データセット間における比較を行う。

4.1 データセット

本研究では、「OPPORTUNITY Activity Recognition Data Set」[14]、「PAMAP2 Dataset」[11]、「CSLab Dataset」[12]の3つのデータセットを用いて分析を行う。以下にデータセットの概要とデータセットの分割方法に関して説明する。

4.1.1 OPPORTUNITY

「OPPORTUNITY Activity Recognition Data Set」は2種類のラベルが付けられており、「Locomotion」と呼ばれる「Stand」「Lie」などユーザの状態を表す4種類の行動クラス(LOC)と、「Gesture」と呼ばれる「Close Door1, Drink Cup」など17の中レベルの行動クラス(GES)がある。本研究ではこの2つを別々のデータセットとして扱い比較をしていく。行動クラスの詳細は表1に示す。

データは指定された行動を順に行う「Drill Session」と、順番が決定していない「ADL Session」の2つがあり、各ユーザはDrill Sessionを1回、ASL Sessionを5回行う。Francisco[14]らの先行研究を参考にして、被験者2-3の3回目のADL SessionをValidationデータに、ユーザ2-3の4-5回目のADL SessionのデータをTestデータに、残りをTrainingデータとして使用する。センサは右腕のIMUの加速度3軸^{*1}と、左足首の加速度3軸^{*2}の合計6軸を使用する。

またセンサの装着位置に関する違いを比較するため、LOCに関してセンサの設置位置を変更した4つのデータセットを追加で作成した。

- (RH-LS: 右手 (RLA) と左足 (R-SHOE), LOC と同じ)
- LH-RS: 左手 (RLA) と右足 (R-SHOE)
- LH-LS: 左手 (LLA) と左足 (L-SHOE)
- RLH: 右手 (RLA) と左手 (LLA)
- RLS: 右足 (R-SHOE) と左足 (L-SHOE)

センサ位置の転移に関してはRH-LS (LOC) を転移先データセットとし、他の5種類のデータセットを転移元データセットとして用いた。

4.1.2 PAMAP2

「PAMAP2 dataset」(PAM)は、9名の被験者に3つ

のIMUと心拍センサを装着し、家事やエクササイズをしている際のデータが10時間にわたって記録されている。行動クラスは移動動作から家事動作など多岐にわたる行動クラスが17種あり、詳細は表1に示す。

Münzner [9]の研究を参考にし、被験者2をTestデータに、被験者3-4をValidationデータに、残りのユーザをTrainingデータとして使用する。またセンサは、手と足首に装着した加速度センサ(16G)のx,y,zの3軸、合計6つのセンサを使用する。

4.1.3 CSLab

「CSLab」は合計62名の被験者の右手・左手・右足・腰の4箇所に加速度センサを装着して、行動のデータが記録されている。行動クラスは14種類あり、詳細は表1に示す。

本研究においては、被験者42-47の7名分をValidationデータ、被験者49-61の13名分をTestデータ、残りの被験者を学習データとして使用する。またセンサとして右足と右足の3軸加速度の合計6つのセンサを使用する。

4.2 データ前処理とニューラルネットワークの学習

センサデータの前処理として、加速度データはデータセット共通に最大+3G、最小-3Gで正規化を行う。またサンプリングレートはおよそ30Hzに揃えるため、PAMに関しては30Hz^{*3}にダウンサンプリングを行う。ニューラルネットワークへの入力、各データセット30Hzの加速度信号をウィンドウサイズ1sec(=30pts, 50% overlap)として切り出したものを1サンプルとし、ラベルはウィンドウ内最後の時刻に対応するラベルを採用した。また学習・検証に使用するサンプルからは、何もしていない時のサンプル(Nullクラス)は除外して学習を行う。

ネットワークの学習は、Adam optimizerを用いてバッチサイズは1024、学習エポック数は300とした。学習係数は $lr = 10^{-4}$ である。モデルの評価は各行動クラスに対するスコアにサンプル数によって重み付き平均をとったWeighted Precision, Recall, F1-measureを用いる。それぞれのモデルを学習した結果とテストサンプル数を表2と表3に示す。

4.3 データセット間の抽出特徴の比較

上記の3種類のデータセットに対して、第3節で述べた(1)ユニットに強く反応する入力信号の可視化手法および、(2)ユニットと行動ラベルの関係性に関する可視化手法を適用し、データセットの比較を行う。

4.3.1 ユニットが強く反応する入力信号 x^u による可視化を用いた比較

3.2節で提案した、ハイパスフィルタを用いて行動認識

^{*1} IMU_accX_RLA, IMU_accY_RLA, IMU_accZ_RLA,

^{*2} IMU_Body_Ax_L-SHOE, IMU_Body_Ay_L-SHOE, IMU_Body_Az_L-SHOE,

^{*3} 正確には33.3Hz

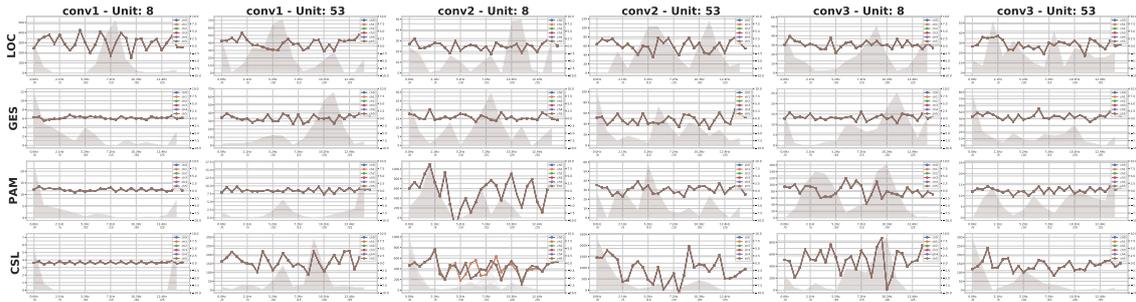


図5 CNNフィルタが強く反応する信号 \mathbf{x}^u の抜粋. 各データセットで学習したモデルの1-3層目の $u = 8, 53$ のユニットに関して可視化を行なった. \mathbf{x}^u は折れ線グラフ (目盛りは右縦軸), \mathbf{x}^u のパワースペクトルは塗りつぶし領域 (目盛りは右縦軸) で可視化した.

モデル中のそれぞれのユニットがもっとも強く反応する信号 \mathbf{x}^u を計算した結果を図5に示す.

本研究で実験に用いた4つのデータセットに共通する内容として周波数成分に着目すると, 1層目の信号は比較的単純な信号が多く, 2層目になると複数のピークがある複雑な信号が多い. そして3層目になるとピークが比較的大きくはっきりとした信号が多くなる. これは1層目では単純な特徴を抽出し, 3層目はある程度行動クラスに特化した特徴を抽出しているため, その行動クラスに特有の周波数が強く出ていると考えられる. またLOCとGESを比較すると, 出力の行動ラベルが異なるだけで学習したデータは同じにも関わらず, 1層目から信号の形が異なるものが多く存在する. CNNの1層目から抽出される特徴は認識する行動クラスに影響されると考えられる.

図6に4つのデータセットから計算した \mathbf{x}^u を, t-SNEを用いて2次元に次元圧縮した結果を示す. 1層目はデータセットが異なっても, \mathbf{x}^u の分布は大きく変わらない. しかし, 深い層になるにつれてデータセットごとに傾向が別れてくる. LOCの3層目の分布は他3つに比べて密集度合いが高い. 一方でCSLは広い範囲に分布している. LOCは行動クラスが4種類と少なく, これに対してCSLは行動クラスが14種類と相対的に多い. 比較的深い層である3層目では一般的な特徴より, データセットの行動クラスに特化した特徴が抽出されていると考えられる. したがって, LOCは行動クラスが少ないため認識に必要な特徴の多様性が低く, \mathbf{x}^u が密集したと考えられる. GESはクラス数が18種類と最も多いが, \mathbf{x}^u の密集度合いが比較的高い. 異なる扉の「Open/Close」など, 行動の内容は非常に似ているものが多いため, 認識に必要な特徴の多様性が低くこのような結果になったと考えられる.

4.3.2 ユニットと行動クラスの関係の可視化を用いた比較

3.3節で提案したユニットと行動クラスの関係性に関する可視化方法を4つのデータセットに用いた結果を基に考察を行う. 個々の可視化の結果はページ数の制約上割愛する. 代わりに各データセットにおいて計算した層ごとのエントロピーの分布とその総和 H^{total} を図7に示す.

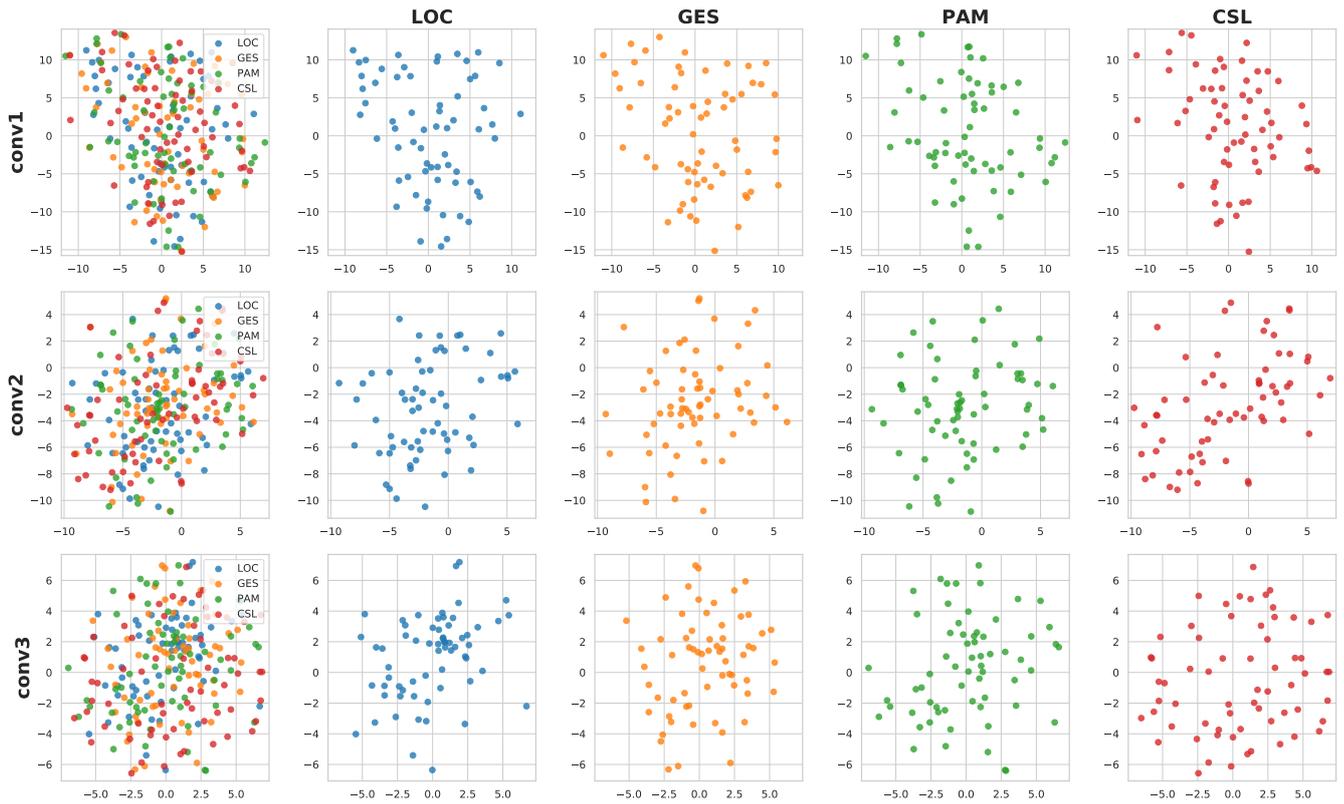
図7[a]からGES, PAM, CSLではエントロピーが高いユニットも多数存在する事がわかる. エントロピーが高くなっているユニットが反応している具体的なラベルをみると, 様々な行動クラスに満遍なく値を出力しているユニットもあるが, いくつかの行動に特に強く反応しているものも多く見られた. GESは「Drink Cup」, PAMとCSLは「Walk」や「Run」などの行動に対してどの層においても多数のユニットが強く反応していた. これらの行動は, サンプル数が多くかつ大きな体の動作が伴うという共通な特徴がある. したがってニューラルネットワークはサンプル数が多い行動クラスを重点的に学習していると考えられる. 一方でLOCは他のデータセットに比べて, エントロピーが高いユニットが少ない. これは行動クラスが4クラスと他より少ないことと, 1層目から特定の行動に強く反応するユニットが多いためである. また, 図7[a]から深い層ほどエントロピーが小さいユニットが増加していることがわかる. 特に2層目から3層目はエントロピーが大きいユニットが大幅に減少している. したがって, 2層目までは一般的な特徴を抽出し, 3層目ではより行動ラベルに特化した特徴に変換されていると考えることができる.

また図7[b]は, センサの装着位置を変えた場合のエントロピーの分布である. 手足の両方を入力しているモデル(LOC, LH-LS, LH-RS)は, 互いに近い分布となっている. 両手を入力とするRLHの1層目は明らかに他とは異なる分布を取り, 1-2層目ではエントロピーが低いユニットが多く存在する. 具体的な行動ラベルを見ると1-2層目は「Stand」と「Sit」に反応が集中している. 他のデータセットでは「Walk」に強く反応するものが多いが, RLHだけはほとんど反応していなかった. RLHは抽出された特徴が他とは明らかに異なると考えられる.

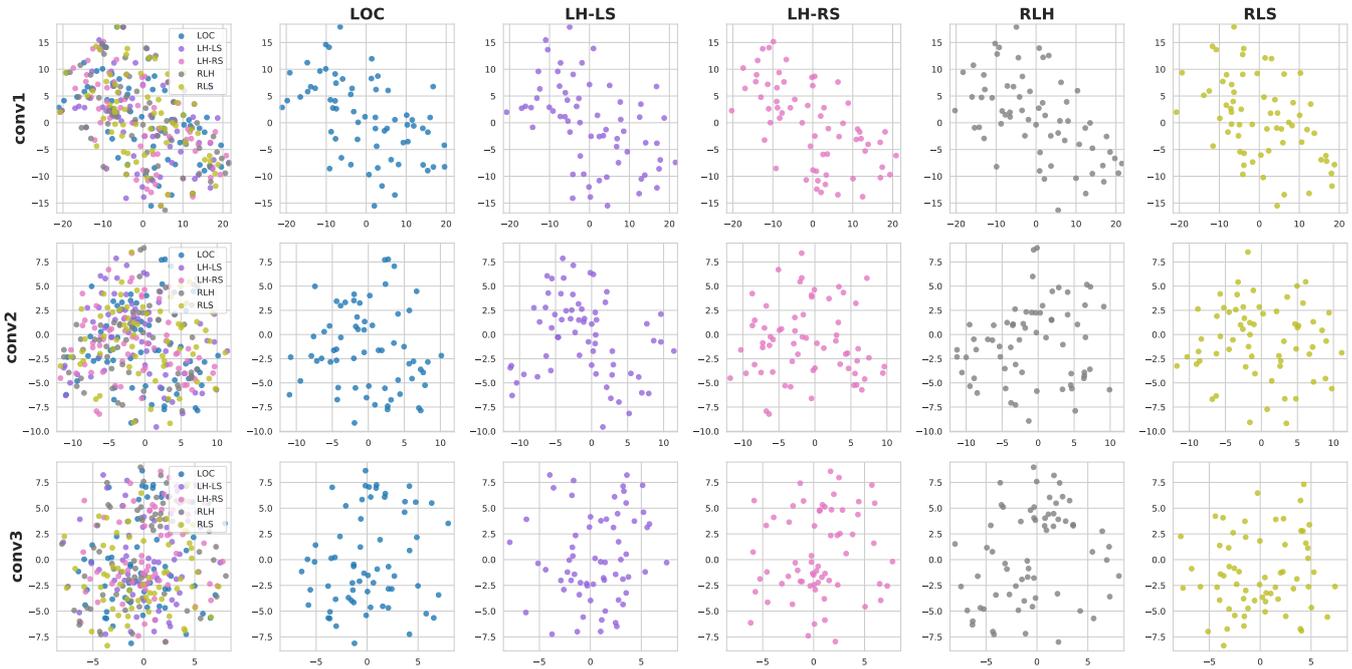
5. 可視化結果に基づく転移学習

5.1 転移学習手法

上記の考察より, ドメイン間・センサの装着位置間に関して転移学習を行い, 認識精度を比較することで前述の考察の有効性を確認する.



[a] ドメイン間

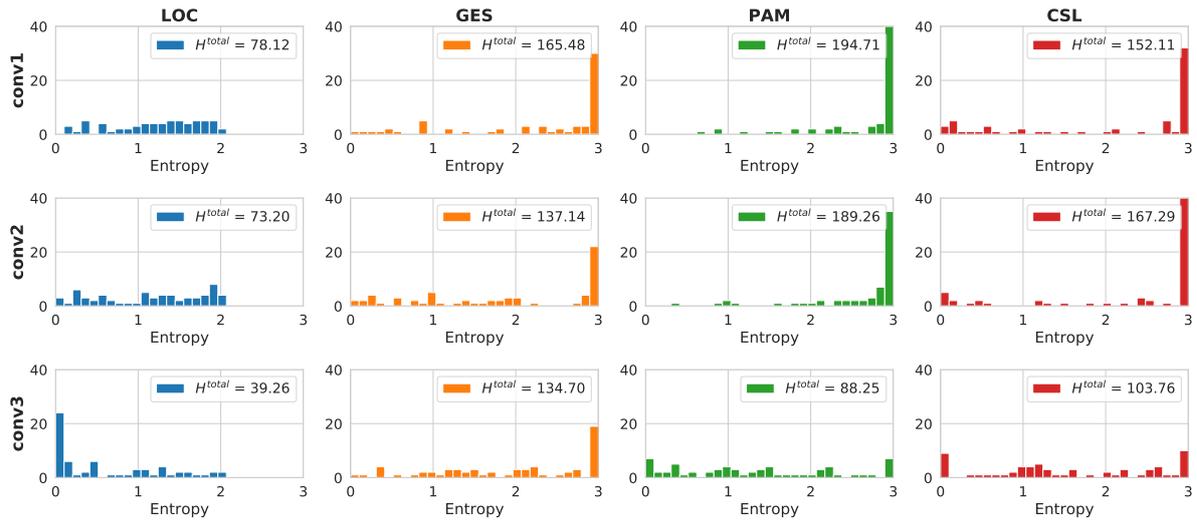


[b] センサ装着位置間

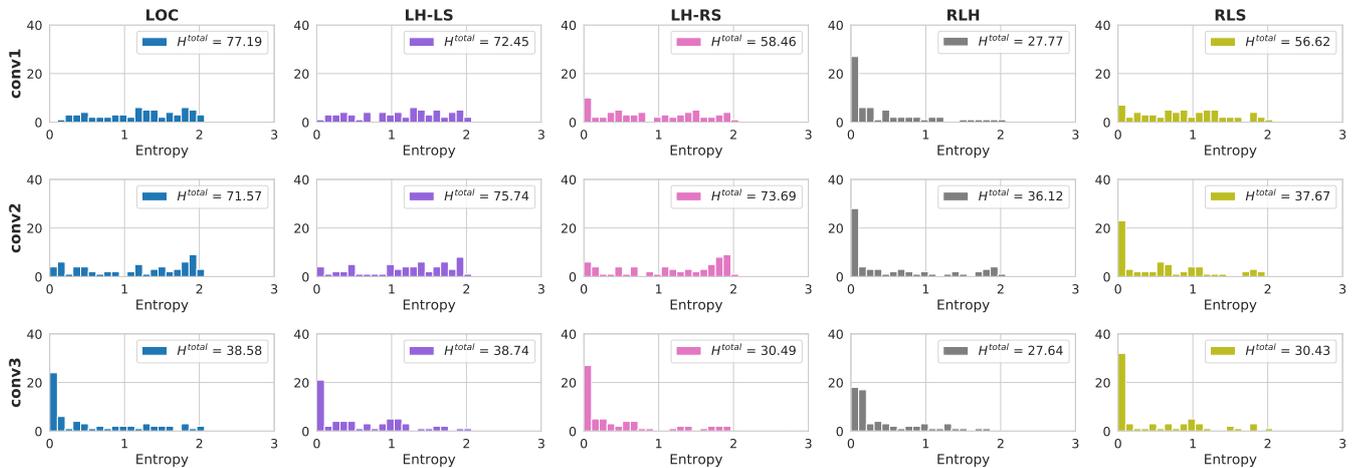
図 6 CNN フィルタが強く反応する信号 x^u の分布。1 列目は 4 つのデータセットの x^u を重ねた結果、2 列目以降は 1 列目の結果から各データセットのデータポイントを抽出した結果である。図 1 に示した通り、conv1 (conv2, conv3) はそれぞれ CNN1 層目 (2 層目, 3 層目) までを転移した結果である。

Morales らの先行研究 [14] に倣い、転移学習の設定を $D_s n D_t (v\%)$ と表す。 D_s, D_t はそれぞれ転移元・転移先のデータセットの短縮表記であり、「LOC, GES, PAM, CSL」

などが当てはまる。また n は転移する CNN の層の数、 v は Fine-tuning の際に転移先の学習データセットから $v\%$ のデータを使用することを示す。本研究では CNN 層の重



[a] ドメイン間



[b] センサ装着位置間

図 7 転移元モデルについて計算した、層ごとのエントロピーの分布とその総和 H^{total}

みの転移に注目するので、 $1 \leq n \leq 3$ である。例えば、転移元モデルとし LOC を用いて、PAM を Training データの 50% を用いて 2 層目まで転移する場合は「LOC2PAM (50%)」と表記する。転移する重みは Fine-tuning の際に、転移元モデルからコピーされた重みも学習し直す方法と、コピーした重みは固定する方法の 2 種類の方法があるが、本研究では固定する方法のみを用いる。

転移先モデル作成は以下の手順で行う。まず、転移元モデルの重みを n 層目まで転移先モデルにコピーする。転移元モデルは、4 節で学習したモデルを用いる。またコピーを行わない層の重みに関しては乱数で初期化を行う。転移先環境の学習データは、あらかじめ学習データ全体からランダムに $v\%$ のサンプルをランダムに抽出しておく。またデータセットの分割は 4 節で用いたものと同じものを用いる。最後に学習率 $l_r = 10^{-4}$ 、 $batchsize = 1024$ を用いて、Adam Optimizer で 20 epoch 学習する。

5.2 評価指標

転移学習の結果の評価には、4.2 節と同様にクラスのサンプル数で重み付き平均した F1-measure を用いる。ただし転移学習の評価に関しては、トレーニングデータの 50% を使用し ($v = 50$)、それぞれの設定において学習を 5 回ずつ行い、その平均値を報告する。

5.3 転移学習の指針

転移学習の方針として、これまでの分析に基づいて転移元のデータセットと転移する層の 2 つを考える。まず、データセット間の転移に着目する。図 6[a] のユニットに最も強く反応する信号 x^u の分布図より、分布の範囲がより広いデータセットを優先して転移元のデータセットとして選ぶ。これは分布範囲が広い方が未知のデータに対応できる可能性が高いからである。したがって他の条件が同じ場合は、CSL \rightarrow PAM \rightarrow GES \rightarrow LOC の優先順位で選ぶ。

次に 3.3 節で提案した 2 つ目の可視化分析の結果を用い

表 4 各データセットにおける、層毎のエントロピーの総和 H^{total} . 転移学習の方針を決めるため、転移元モデルに転移先の学習データを入力した結果である.

転移先	転移元	conv1	conv2	conv3
LOC	GES	29.6	28.1	2.4
	PAM	38.5	30.4	3.3
	CSL	33.7	24.4	11.5
GES	LOC	8.7	7.6	4.8
	PAM	19.0	21.7	0.5
	CSL	26.1	15.0	3.3
PAM	LOC	17.6	13.1	9.7
	GES	85.6	60.9	42.1
	CSL	107.5	80.8	63.5
CSL	LOC	5.4	4.0	0.4
	GES	110.0	86.4	25.9
	PAM	120.8	75.2	44.0
LOC (装着位置)	LH-LS	72.5	75.7	38.7
	LH-RS	58.5	73.7	30.5
	RLH	27.8	36.1	27.6
	RLS	56.6	37.7	30.4

ることを考える。ただし、転移学習を行うにあたって転移元のモデルからは転移先のクラスの推定確率 \hat{y}_i を得ることができない。したがって、転移先ラベルなしデータを入力した際の転移元の行動クラスに対する出力値の分布および各ユニットのエントロピーから転移学習を行う転移元データセットと転移する層を決める。個々の可視化の結果に関してはスペースの都合上割愛する。ただし層毎のエントロピーの総和 H^{total} に関しては表 4 にまとめた。転移元の行動クラスに対して、出力値の分布の偏りが大きい場合は転移元モデルに適さないと考えられる。また、浅い層ではユニットのエントロピーが高い方が行動ラベルに依存しない特徴を抽出できるため転移学習に適している。一方で深い層ではユニットではある程度行動クラスに特化している方が学習が必要なパラメータを減らすことができるので転移学習に望ましい。

上記の判断基準に加え、転移する層に決め手となる判断基準がない場合は可能な限り深い層まで転移を行う。これは学習パラメータを減らすことで、転移先の学習データで過学習することを抑制するためである。

以上を踏まえて、4 種類それぞれを転移先環境とした場合の転移学習の方針を考察し、転移先環境のデータを 50% 使用し CNN 層の重みを転移する設定において、精度が高くなると考えられる 3 候補を示す。

5.3.1 転移先: LOC

ユニットと反応する行動クラスの間接性を見ると、どのデータセットも LOC のデータを入力した場合の出力が非常に偏っていた。また表 4 から分かるように、GES・PAM の 3 層目のエントロピー H^{total} は非常に小さかった。GES と PAM の 3 層目はそれぞれ「Drink cup, Lying」にしか

反応しないため転移学習には適さないと考えられる。一方で CSL は反応している行動クラスが複数あった。したがって 3 層目を転移する場合は CSL, 2 層目を転移する場合は PAM または CSL が適切だと考えられる。また 1 層目と 2 層目はエントロピーを計算すると大きな変化はないため、学習パラメータ量を減らせる 2 層目を使う方が良いと考えられる。

したがって候補は「CSL3LOC, CSL2LOC, PAM2LOC」である。

5.3.2 転移先: GES

ユニットの出力と行動クラスの間接性について可視化を行うと、PAM, CSL の 2,3 層目はユニットの反応が非常に偏っていた。表 4 に示したように、エントロピーの合計 H^{total} は全体的に小さいが、3 層目は特に小さい値になっていた。しかし 3 層目は特定の行動に偏ってエントロピーが小さくなっていたので、転移学習には適さない。

したがって 3 層目の転移を除外し、他は前述のルールに従い、「CSL2GES, PAM2GES, LOC2GES」が候補になる。

5.3.3 転移先: PAM

ユニットの出力と行動クラスの間接性について可視化を行うと、LOC と GES を転移先とした場合と異なり、2-3 種類の転移元の行動ラベルが反応した。表 4 から GES・CSL のエントロピー H^{total} も非常に大きな値を取っていることが分かる。しかしユニットごとのエントロピーを計算すると、3 層目においてもエントロピーが低いものから高いものまで広く分布しており、転移先環境においても特定の行動に強く反応するものと汎用的な特徴に反応するものが両方存在していると考えられる。ただし、PAM は「Walk」などの移動行動が含まれるが、LOC の「Walk」はほとんど反応していなかった。したがって LOC は転移元として相対的に適していないと考えられる。

以上の考察から、「CSL3PAM, GES3PAM, CSL2PAM」が候補となる。

5.3.4 転移先: CSLab

転移先環境を PAM とした場合と同じ議論ができる。したがって GES または PAM を転移元モデルとして 3 層目までを転移すると良いと考えられるので、「PAM3CSL, GES3CSL, PAM2CSL」が候補となる。

5.3.5 転移先: LOC (装着位置の転移)

次に LOC におけるセンサ装着位置間の転移に注目する。図 6[b] より、両手のセンサを用いる RLH は 3 層目の中央右側部分が LOC と大きく異なっていた。これは足の動きに関する特徴を学習したユニットと考えられる。分布が大きく異なるので RLH は転移元から除外する。また両足のセンサの RLS は他に比べてエントロピーが低いものが多いが、強く反応する行動クラスを見ると「Walk, Sit, Lie」に分かれていた。これは多くの行動に特化したユニットの

集まりであることを意味するので、転移元に適していると考えられる。残りの手足の両方を含む LH-RS・LH-LS は反応が強い行動クラスが偏りを見ると、どちらとも「Sit」または「Stand」に集中しており、優先度を決定する判断材料とはならない。しかし表 4 に示す通り、RLS の 2 層目のエントロピーの総和は、他の 2 層目の結果に対して明らかに小さい。RLS の 2 層目が行動に特化しすぎている可能性があるため、RLS の 2 層目までより LH-RS・LH-LA の 3 層目を優先して使うこととする。

したがって、「LH-LS3LOC, RLS3LOC, LH-RS3LOC」が候補となる。

5.4 転移学習結果

転移学習を全ての設定において行い、F1 値が高かった上位 3 件の設定を表 5 示す。また結果の詳細は図 8-12 に示す。参考のため今回対象としなかった、4-5 層目の全結合層と、転移先データの利用率 $v = 100\%$, 75% の結果も合わせて示す。

転移先の学習データが 50% の場合、GES 以外は転移学習によって転移先の学習データのみ ($v = 50\%$) で学習した場合の認識精度を超えた。また Münzner [9] らは転移する層が増えるごとに転移先での認識精度が低下することを報告したが、本研究においては、多くの場合で CNN そうに関しては深い層まで転移した方が認識精度が向上する結果となった。使用したデータセットや CNN 層の後でセンサフュージョンを行うモデルの構成が異なるためだと考えられる。

前節で予測した転移学習の設定は、LOC・PAM・CSL は 3 候補中 2 件が上位 2 件に入った。したがって提案した可視化を用いた考察は有効であると考えられる。しかし転移先が GES の場合は、3 層目まで転移する設定が高い F1 値となった。GES はそもそもの認識が難しいため、特徴抽出部分を固定して全結合層を効率よく学習する事が認識精度の向上につながったと考えられる。

センサの装着位置の転移学習においては、予想通り RLH 以外は 3 層目まで使用する方法が高い F1 値となった。F1 値の上位 3 件は表 5 の通りになっているが、3 層目まで転移した結果同士で認識精度に大きな差は無い。また LH-RS, LH-LS の結果から、行動が左右対称であるという本研究の設定の設定においては、左右の転移は抽出される特徴が似ているため、転移が容易であるという事が確認できた。

6. 結論

本研究では行動認識分野での転移学習の有効利用を促進するため、行動認識モデルの分析と転移学習を行う方針に関して考察を行なった。モデルの分析では、まずニューラルネットワーク中の各ユニットの出力が最も大きくなる入力信号を勾配上昇法で求めた。単純に勾配上昇法を適用す

ると高周波成分が多く含まれる信号が生成される。人間の手足の動作の周波数は比較的低いため、ハイパスフィルタを用いた正則化項を提案し、より行動認識モデルに適した可視化を行なった。また各ユニットの出力を認識モデルの推定結果を用いて行動クラスと結びつける事で、注目したユニットがどの行動クラスの推定に寄与しているかを可視化した。これらの可視化方法を用いて異なるデータセットを学習したモデルを比較した。可視化結果から 1-2 層目は比較的一般的な特徴を抽出しており、3 層目はある行動ラベルに特化した特徴を抽出している事がわかった。この分析結果を基に、(1) ユニットの出力を最大化する信号の分布と (2) 転移先のデータを入力した際のユニットの出力と行動クラスの関係性の 2 点に基づいて、転移学習を行う方針を考察し、実際に転移学習を行い分析が有効であることを確かめた。

今回の分析を用いる事でより精度が高い転移学習の設定を推定する事はできたが、転移学習の方針の決定は定量的ではなかった。転移学習の方針を決定する際に参考にした視覚的情報を定量化することは、より複数の転移元モデルを比較する場合や、多数の層を持つネットワークに提案手法を応用する上で必要である。また、転移学習によって転移元だけで学習したモデルより大きく認識精度が向上する事例は少なかった。転移学習をより効果的に行うモデルの構築や、より良い転移学習の方法だけでなくどの程度認識精度が向上する可能性があるか推定することも重要である。今後はこれらの課題に取り組んでいきたい。

謝辞

本研究の一部は JST CREST JPMJCR15E2, JSPS 科研費 JP16H06539, JP17H04679 の助成を受けて行われたものです。データセットの一部は日本電信電話株式会社 コミュニケーション科学基礎研究所から提供を受けたものです。

参考文献

- [1] Yosinski, J., Clune, J., Bengio, Y. and Lipson, H.: How transferable are features in deep neural networks?, *Advances in neural information processing systems*, pp. 3320–3328 (2014).
- [2] Mahendran, A. and Vedaldi, A.: Understanding deep image representations by inverting them, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5188–5196 (2015).
- [3] Yosinski, J., Clune, J., Nguyen, A., Fuchs, T. and Lipson, H.: Understanding neural networks through deep visualization, *arXiv preprint arXiv:1506.06579* (2015).
- [4] Ranjan, R., Patel, V. M. and Chellappa, R.: Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 41, No. 1, pp. 121–135 (2019).

表 5 転移学習の結果. 各転移先環境において F1 値が高い上位 3 件ずつを示す. 太字は 5.3 節で候補とした転移学習の設定である.

Source	LOC		GES		PAM		CSL		LOC (装着位置)	
	手法	F1	手法	F1	手法	F1	手法	F1	手法	F1
Top1	PAM3LOC	0.863	PAM3GES	0.398	CSL3PAM	0.832	PAM3CSL	0.827	RLS3LOC	0.861
Top2	CSL2LOC	0.860	CSL2GES	0.394	LOC3PAM	0.810	LOC2CSL	0.824	LH-RS3LOC	0.860
Top3	CSL3LOC	0.860	CSL3GES	0.392	GES3PAM	0.809	PAM2CSL	0.824	LH-LS3LOC	0.860

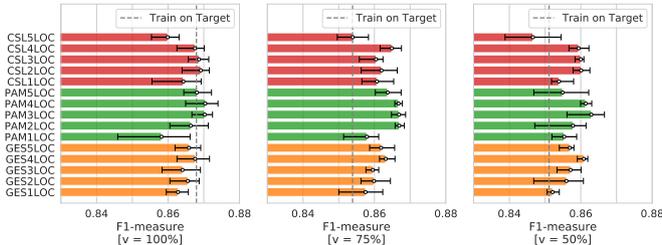


図 8 転移学習の結果: 転移先 LOC

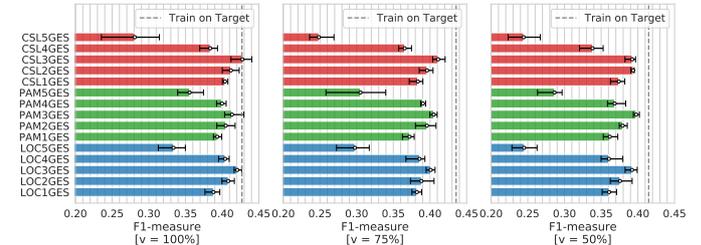


図 9 転移学習の結果: 転移先 GES

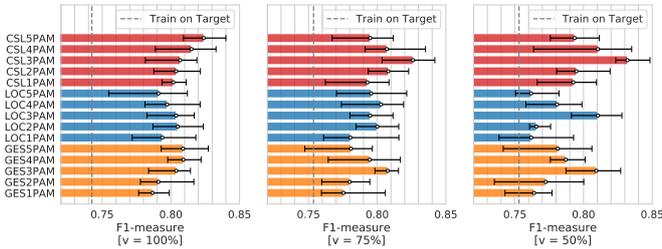


図 10 転移学習の結果: 転移先 PAM

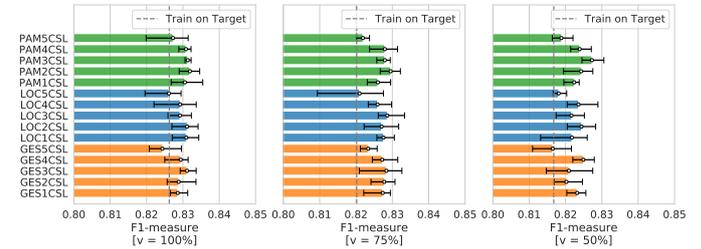


図 11 転移学習の結果: 転移先 CSL

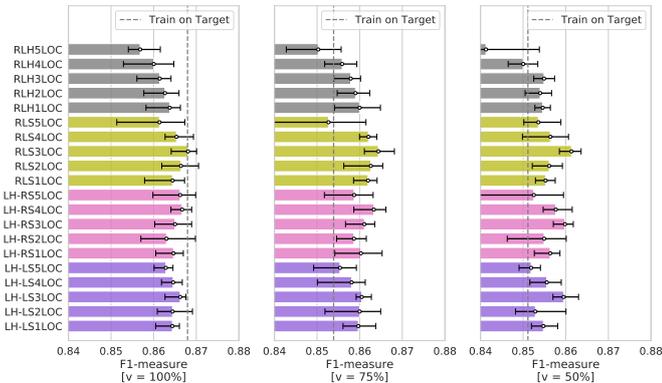


図 12 転移学習の結果: LOC (センサ装着位置間の転移)

[5] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J. and Zehan Wang, W. S.: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, *arXiv preprint arXiv* (2016).

[6] Hatori, J., Kikuchi, Y., Kobayashi, S., Takahashi, K., Tsuboi, Y., Unno, Y., Ko, W. and Tan, J.: Interactively picking real-world objects with unconstrained spoken language instructions, *2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 3774–3781 (2018).

[7] Krizhevsky, A., Sutskever, I. and Hinton, G. E.: ImageNet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp. 1097–1105 (2012).

[8] Ordóñez, F. and Roggen, D.: Deep convolutional and

lstm recurrent neural networks for multimodal wearable activity recognition, *Sensors*, Vol. 16, No. 1, p. 115 (2016).

[9] Münzner, S., Schmidt, P., Reiss, A., Hanselmann, M., Stiefelhagen, R. and Dürichen, R.: CNN-based sensor fusion techniques for multimodal human activity recognition, *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pp. 158–165 (2017).

[10] Roggen, D., Calatroni, A., Rossi, M., Holleczeck, T., Förster, K., Tröster, G., Lukowicz, P., Bannach, D., Pirkel, G., Ferscha, A. et al.: Collecting complex activity datasets in highly rich networked sensor environments, *2010 Seventh international conference on networked sensing systems (INSS)*, IEEE, pp. 233–240 (2010).

[11] Reiss, A. and Stricker, D.: Introducing a new benchmarked dataset for activity monitoring, *2012 16th In-*

- ternational Symposium on Wearable Computers*, IEEE, pp. 108–109 (2012).
- [12] Maekawa, T. and Watanabe, S.: Unsupervised activity recognition with user’s physical characteristics data, *2011 15th Annual International Symposium on Wearable Computers*, IEEE, pp. 89–96 (2011).
- [13] Bachlin, M., Roggen, D., Troster, G., Plotnik, M., Inbar, N., Meidan, I., Herman, T., Brozgol, M., Shaviv, E., Giladi, N. et al.: Potentials of enhanced context awareness in wearable assistants for Parkinson’s disease patients with the freezing of gait syndrome, *2009 International Symposium on Wearable Computers*, IEEE, pp. 123–130 (2009).
- [14] Morales, F. J. O. and Roggen, D.: Deep convolutional feature transfer across mobile activity recognition domains, sensor modalities and locations, *Proceedings of the 2016 ACM International Symposium on Wearable Computers*, pp. 92–99 (2016).
- [15] Bouten, C. V., Koekkoek, K. T., Verduin, M., Kodde, R. and Janssen, J. D.: A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity, *IEEE transactions on biomedical engineering*, Vol. 44, No. 3, pp. 136–147 (1997).
- [16] Maekawa, T., Yanagisawa, Y., Kishino, Y., Ishiguro, K., Kamei, K., Sakurai, Y. and Okadome, T.: Object-based activity recognition with heterogeneous sensors on wrist, *International Conference on Pervasive Computing*, Springer, pp. 246–264 (2010).
- [17] Saks, S.: *Theory of the integral. 2. ed. English translation by L. C. Young. With two additional notes by S. Banach.*, Vol. 7, PWN - Państwowe Wydawnictwo Naukowe, Warszawa (1937).