Vol.2021-MBL-101 No.3
Vol.2021-ITS-87 No.3
2021/12/7

加速度センサデータを用いた深層学習モデルの教師なし表現学習手法

武仲 紘輝[1,a)]　長谷川 達人[1,b)]

概要：深層学習は加速度センサデータを用いた行動認識の推定精度向上に寄与している．深層学習を用いた行動認識では一般的に行動ラベルが付いた加速度センサデータを訓練データとして教師ありで学習が行われる．近年ではモバイル端末の普及により大量の加速度センサデータを容易に入手できるが，計測されたデータは行動ラベルのない場合が多い．大量の加速度センサデータに行動ラベルを目視でアノテーションする作業は膨大な時間や人手を必要とする問題点がある．そのため行動ラベルのない加速度センサデータを用いて，深層学習モデルに行動認識の特徴表現を獲得させる手法が求められる．そこで本研究では，画像認識分野で提案された Instance Discrimination と Feature Independent Softmax を用いた教師なし表現学習手法をベースに，Segment Discrimination と Autoencoder を組み合わせた手法を提案し，表現学習の精度検証を行った結果を報告する．

1. はじめに

加速度やジャイロなどのセンサを用いて計測したデータからユーザがどのような行動を行っていたのかを予測する行動認識が盛んに研究されている．センシングによる行動認識技術は，高齢者の見守りシステムやヘルスケアアプリケーションなどへの応用が見込める．これらの応用を考えると，認識結果は様々な意思決定に利用されることから，より高精度な認識技術が求められている．

近年の行動認識は深層学習を用いて実現されることが多い [1], [2], [3]．従来は，センサデータからの特徴抽出を手動で行い，機械学習によって予測モデルを構築していることが多かったが，深層学習ではセンサデータを直接モデルに入力し，特徴表現の自動獲得（表現学習）を行うことで高精度を達成できる．しかし，深層学習モデルが特徴表現を獲得するためには大量の訓練データが必要である．行動認識における，多くの表現学習は教師あり学習によって実現されている．従って，行動認識の訓練データは，モデルへの入力となるセンサデータに対して，対応する行動ラベルを付与したものとなる．一方で，センサデータに対応する行動ラベルを付与するアノテーションは，多大な労力を要するという課題が残っている．例えば，センサを装着した被験者が特定の動作を行う際の動画を撮影しておき，動画から人間が目視で動作を判定し，時刻同期を行ったセンサ値に対して行動ラベルを付与する必要がある．

近年ではスマートフォンやウェアラブルデバイスなどのセンシング端末の普及によりラベルなしセンサデータを大量に計測することが容易になった．大量のラベルなしセンサデータを用いて特徴表現の獲得を行うことができれば，上記の問題点を解決することが可能である．

以上を踏まえ，本研究では画像認識分野で提案された Instance Discrimination (ID) と Feature Independent Softmax (FIS) を用いた教師なし表現学習手法を行動認識に適用する手法を新たに開発する．画像の入力をセンサ波形に変更するシンプルな変更に加えて，センサ波形がシーケンシャルに発生し，人間の行動は時系列連続性があることを考慮し，新たに Segment Discrimination (SD) を提案する．更に，Autoencoder (AE) による Reconstruction Loss を加えることで推定精度の向上を図る．これらの提案手法により，行動認識における教師なし表現学習手法を確立し，行動認識に有用な特徴表現の獲得を目的とする．

本研究の貢献は以下のとおりである．

- 画像認識分野で提案された ID と FIS を用いた表現学習手法を行動認識に適用した．
- センサデータの特性を考慮し ID を改良した SD を新たに提案し，推定精度が向上することを実験により明らかにした．
- AE による Reconstruction Loss を追加で導入することで，より有用な特徴表現が得られることを示した．

[1] 福井大学大学院工学研究科
Graduate School for Engineering, University of Fukui
[a)] ktakenak@u-fukui.ac.jp
[b)] t-hase@u-fukui.ac.jp

2. 関連研究

2.1 行動認識と深層学習

従来の行動認識は, 多くの研究がセンサデータを入力とし行動ラベルを出力とする教師あり学習により実現されている．特に深層学習を用いた事例では, Convolutional Neural Network (CNN) や Recurrent Neural Network (RNN) を用いた取り組みが多い．Zeng ら [1] は CNN を用いてモバイル端末により計測されたデータの行動認識に取り組んでいる．Murad ら [2] は複数の LSTM を重ねた Deep RNN (DRNN) を用いて unidirectional DRNN，bidirectional DRNN を提案し，複数の行動認識データセットで行動認識の有効性を実証している．

シンプルな CNN を用いながらも，入力に工夫を行っている事例もある．Mahmud ら [3] は入力する加速度センサデータに Gramian Anguler Filed などの複数の変換手法を適用し，変換したデータそれぞれに小規模な CNN を構築し，複数の段階に分け訓練を行う訓練手法を提案している．最初の段階はそれぞれの CNN に対して訓練を行い，次に訓練した CNN を組み合わせたアンサンブルモデルを構築し訓練する．行動認識推定時は構築したアンサンブルモデルを用いる．提案した訓練手法により行動認識精度が向上することを示している．

多くの深層学習モデルが提案されている一方で，大部分が教師あり学習を前提としているため，特に近年の深く広いモデルを訓練するためには，大量のセンサデータと対応する行動ラベルが必要である．

2.2 行動認識における教師なし特徴表現学習

大規模データセットのアノテーション作業は時間や人手を必要とし困難なためラベルを用いずにセンサから行動認識に有用な特徴表現を獲得する手法が研究されている．Saeed ら [4] はセンサデータにデータ拡張を適用し，適用したデータ拡張の種類をラベルとし訓練する手法を提案している．これにより訓練したモデルに少量のラベルありデータセットを用いて追加訓練することで行動認識に取り組んでいる．Haresamudram ら [5] は Contrastive Predictive Coding (CPC)[6] と呼ばれる自己教師あり学習手法を行動認識に応用している．CPC は，入力となるセンサデータを深層学習モデルで特徴マップに変換し，未来の時刻の特徴マップを予測するように訓練する Contrastive Learning 手法である．Sheng ら [7] は AE を利用した行動認識の教師なし表現学習に取り組んでいる．復元誤差のほかに時間的近傍の波形や特徴の類似した波形と復元した波形が一致するような損失関数を追加することで行動認識に有用な特徴表現を獲得することを期待してる．波形の特徴は平均や標準偏差などの基本的な統計量を用いて手動で求めている．これらの研究は時系列データの並びや計測された時刻の近い波形は似た波形となっている特性を訓練に活用することで行動認識に有用な特徴表現の獲得を期待している．本研究では特徴マップを活用した表現学習手法に焦点を当て，提案する SD が加速度センサデータの特性を活用し，行動認識に有用な特徴表現を獲得することを期待する．

2.3 画像認識における教師なし特徴表現学習

画像認識分野でも教師なし表現学習の取り組みがある．Wu ら [8] は訓練データの一つのデータを一つのクラスとみなし訓練を行う ID を提案している．通常，深層学習モデルは C クラス分類を行うときモデルの出力の次元数を C とする．しかし ID ではデータ数 N をクラス数とするため，一般的なモデルを構築する場合，データ数が大きければ大きいほど出力次元数も増加するという問題がある．この問題に対してメモリバンクを利用した Nonparametric Softmax Classifier を提案し，モデルの出力する次元数が増加する問題を解決している．Gansbeke ら [9] は 2 つの深層学習モデル g, f を用いて 3 段階に分けた訓練を行う手法を提案している．提案手法は，まずはじめに，ID などを用いてモデル g に表現学習を行う．そして，事前訓練したモデル g から得られる特徴マップに対して k 近傍法を適用し，クラスタリング結果を入力データに対応する疑似ラベルとする．疑似ラベルを用いて提案された損失関数 SCAN-loss で一方のモデル f の訓練を行う．最後に，モデル f の予測結果を疑似ラベルとしてモデル f の訓練を行う．得られたモデル f を最終的な予測モデルとしている．Tao ら [10] は非直行成分が非ゼロとなることを許容した特徴マップに対する直行制約である FIS を提案し，ID と組み合わせることで画像に存在する小さな特徴も反映する特徴表現を獲得することに取り組んでいる．

これらの手法はモデルの構造に依存しておらず，加速度センサデータの教師なし表現学習に適用しやすい手法である．しかし画像に特化した手法であるため加速度センサデータに適用するためには何らかの工夫が必要である．

3. 提案手法

3.1 問題設定

本研究では，行動認識における教師なし表現学習手法を開発するため，以下の問題設定を前提として，以降の説明を行う．計測されたデータは予めセグメントで分けられていることを想定している（**図 1**）．セグメントごとにスライディングウィンドウ方式を用いて入力データを作成する．入力は加速度等のセンサ波形データを扱うため，ある入力 $\boldsymbol{s}$ は，$\boldsymbol{s} = [\boldsymbol{x}_i, \boldsymbol{x}_{i+1}, \ldots, \boldsymbol{x}_{i+L}]$ の書式となる．ここで，i は時系列のインデックス番号を，$\boldsymbol{x}_i$ は i におけるセンサ値，L は 1 回の入力で扱う系列長（ウィンドウサイズ）を意味する．$\boldsymbol{x}$ は一般的には x 軸, y 軸, z 軸の 3ch データであることが多い．これに対して行動ラベルデータは付与され

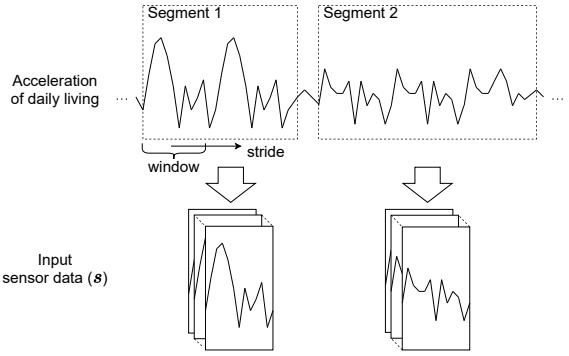

図 1 入力する加速度センサデータ．計測されたデータがセグメント分けされている．入力データはセグメント単位でスライディングウィンドウ方式を用いて作成する．

Fig. 1 Input sensor data. Acceleration of daily living is divided into segments. Input sensor data is created in segments using the sliding window method.

ていないものとし，N 件のセンサ系列を持つラベルなしセンサデータセット $D_{unsup} = \{\boldsymbol{s}_i\}_{i=1}^{N}$ を用いて教師なし表現学習を行う．

次に，モデルの性能評価を行うためには，データ数を M として少量のラベル付きデータセット $D_{sup} = \{(\boldsymbol{s}_i, y_i)\}_{i=1}^{M}$ を用いて追加訓練を行う．ここで，$(\boldsymbol{s}_i, y_i)$ はセンサデータと対応する行動ラベルの組みを表しており $\boldsymbol{s}_i$ は入力するセンサデータ，y_i は入力 $\boldsymbol{s}_i$ に対応する行動ラベルを表している．データセットのデータ数の関係は $M \ll N$ である．また，精度検証のために検証用データセット D_{test} を D_{sup} とは別に準備する．

3.2 提案するモデルの概要

提案する教師なし表現学習手法のモデル構造と訓練手順を**図 2** に示す．実線の矢印はデータの流れを表し，破線の矢印は損失関数への入力を表す．モデル構造自体は AE のシンプルな Encoder-decoder 構造を採用している．Encoder は特徴抽出器として動作する CNN であり，入力したセンサデータを特徴マップ（feature maps）に変換する役割を担う．Decoder は Reconstruction Loss（L_{rec}）を算出するために，特徴マップから入力データを復元する役割を担う．更に，特徴マップを用いて，FIS による損失（L_F）と，SD による損失（L_{SD}）を算出する．L_{rec} はモデル全体の訓練に利用し，L_F, L_{SD} は Encoder の訓練に利用する．各損失関数の詳細は後述する．

教師なし表現学習を行ったモデルを行動認識に活用するには 2 つの方法がある．1 つは，少量のラベル付きデータセット D_{sup} が与えられる環境を想定し，学習済みの Encoder に対して D_{sup} で追加訓練を行う手法（以降，Fine-tuning; FT とする）である．もう 1 つは，学習済みの Encoder が出力する特徴マップから，クラスタリングによって行動を予測する教師なしの行動認識手法（以降，

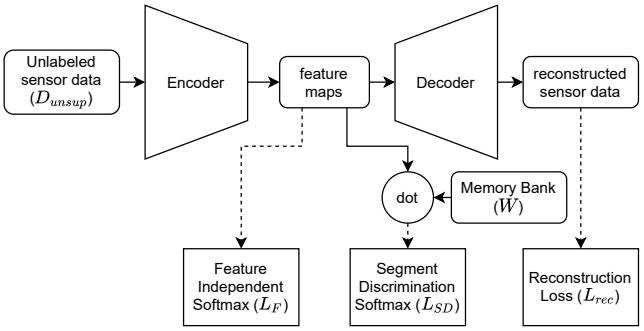

図 2 提案するモデルの概要．実線の矢印はデータの流れを表し，破線の矢印は損失関数への入力を表す．

Fig. 2 Overview of our method. Solid line and dashed line denote data flow and inputs to loss function respectively.

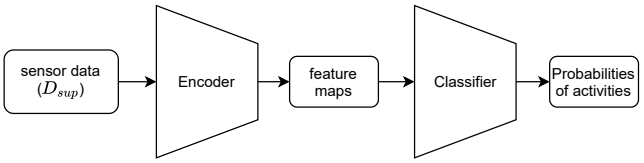

図 3 少量の教師ありデータを用いた追加訓練による行動認識モデル訓練時のモデル構造．図 2 で訓練した Encoder の末尾に新たに Classifier を埋め込んだモデル構造である．

Fig. 3 Model architecture during training of an activity recognition model by additional training using a small amount of supervised data. The new classifier is embedded at the end of the encoder trained in Fig. 2.

Clustering; Clst とする）である．

FT により行動認識を実現する際のモデル構造を**図 3** に示す．FT では学習済みの Encoder に対して，新たに全結合層で構成される Classifier を接続し，D_{sup} を用いて Encoder と Classifier を追加訓練する．Encoder は D_{unsup} により事前訓練されていることから，D_{sup} のような少数のデータでも高精度な行動認識モデルが実現可能となる．

Clst による行動認識は，学習済みの Encoder が出力する特徴マップを k-means 法のような何らかのクラスタリング手法によりグループ分けすることで実現する．このとき，教師データを全く使わないことから，歩行や走行と言った具体的な行動種別の予測までは実現できないものの，何らかの行動単位でクラスタを構築することが可能となる．

3.2.1 Feature Independent Softmax

FIS は Tao ら [10] により提案された手法であり，特徴マップに対する直交制約である．通常の直交制約との違いは FIS では Softmax 関数を利用し非直交成分で非ゼロとなることを許容しており制約が緩和されている点である．Tao らは FIS を用いることで非直行成分で非ゼロを許容しない厳密な直行制約と比べて安定した訓練を行えることを実証している．データ数 N のラベルなしデータセット $D_{unsup} = \{\boldsymbol{s}_i\}_{i=1}^{N}$ から Encoder を用いて求めた i 番目の特徴マップを $\boldsymbol{v}_i$ とする．バッチサイズ b として求め

た特徴マップは $V = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_b]$ と表せる．このとき $F = V^\top$ を考える．i 番目の特徴の b 次元ベクトルを $\boldsymbol{f}_i$ と表すと $F = [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_d]$ と表せられる．FIS は $\boldsymbol{f}_i$ を利用し，損失関数は次のように表せる．

$$L_F = -\sum_{i=1}^{d} \log \frac{\exp(\boldsymbol{f}_i^\top \boldsymbol{f}_i)}{\sum_{j=1}^{d} \exp(\boldsymbol{f}_j^\top \boldsymbol{f}_i)} \tag{1}$$

FIS は損失関数 L_F を最小化するように訓練することで，特徴マップの各要素が互いに独立な情報を得るような制約項の役割を持つ．

3.2.2 Instance Discrimination

Wu ら [8] の提案した ID は一つのデータ（インスタンス）を一つのクラスとみなして訓練を行う教師なし表現学習手法である．ID はメモリバンクを利用した Nonparametric Softmax Classifier を用いて訓練を行う．ID の訓練手法は，はじめに訓練データ $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N$ から特徴マップ $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N$ を求める．特徴マップの次元数 d とするとメモリバンク W は $W \in \mathbb{R}^{N \times d}$ の行列であり，メモリバンクを用いて $\boldsymbol{v}'_i = W\boldsymbol{v}_i$ を計算することで i 番目の特徴マップ $\boldsymbol{v}$ から N 次元の特徴マップ $\boldsymbol{v}'_i$ を求める．このとき，ある入力 $\boldsymbol{s}$ に対応した特徴マップ $\boldsymbol{v}'$ が i 番目のデータの特徴マップである確率を Softmax 関数と温度パラメータ τ を用いて

$$P(i|\boldsymbol{v}') = \frac{\exp(\boldsymbol{v}'_i{}^\top \boldsymbol{v}'/\tau)}{\sum_{j=1}^{N} \exp(\boldsymbol{v}'_j{}^\top \boldsymbol{v}'/\tau)} \tag{2}$$

と表す．ここで $\boldsymbol{v}'_i$ は i 番目のデータの特徴マップを表しており $||\boldsymbol{v}'|| = 1$ となるように正規化している．次に $\prod_{i=1}^{N} P(i|\boldsymbol{v}'_i)$ を最大化すればよいため ID の損失関数は

$$\begin{aligned} L_{ID} &= -\sum_{i=1}^{N} \log P(i|\boldsymbol{v}'_i) \\ &= -\sum_{i=1}^{N} \log \frac{\exp(\boldsymbol{v}'_i{}^\top \boldsymbol{v}'/\tau)}{\sum_{j=1}^{N} \exp(\boldsymbol{v}'_j{}^\top \boldsymbol{v}'/\tau)} \end{aligned} \tag{3}$$

となり，損失関数 L_{ID} を最小化するように訓練を行う．

ID は訓練データ $\boldsymbol{s}_i, \boldsymbol{s}_j$ に対応するそれぞれの特徴マップ $\boldsymbol{v}'_i, \boldsymbol{v}'_j$ に対して内積を用いた類似度を計算し，$i = j$ のとき最大，$i \neq j$ のとき最小となるような類似度を求める役割がある．したがって ID によってインスタンス間の差異を特徴マップとして獲得すると考えられる．

3.2.3 Segment Discrimination

ID はすべてのインスタンスが独立であることを前提とした手法である．一方，行動認識に適用することを考えた場合，入力となるセンサ値は時系列にしたがって連続的に観測されるため，時系列的に近い情報を独立として扱うことが表現学習に悪影響を及ぼす可能性がある．そこで，我々は時系列的な近傍情報に関する ID の制約を緩和する手法として SD を新たに提案する．

提案する SD は同一のセグメントから取得したデータを一つのクラスと見なすという点で ID と異なる．本提案手法を利用するうえでの制約としてラベルのない何らかのセグメントで区切られた加速度センサデータを利用する．セグメントの数を K とし，ある加速度センサデータ $\boldsymbol{s}$ に対応する特徴マップを $\boldsymbol{v}$ とする．そして i 番目のクラスである特徴マップを $\boldsymbol{v}_i$ としたとき，ID と同様にメモリバンクを利用して K 次元の特徴マップ $\boldsymbol{v}'_i$ を求める．このとき温度パラメータ τ を用いた Softmax 関数を用いて $\boldsymbol{s}$ が i 番目のクラスである確率は

$$P(i|\boldsymbol{v}') = \frac{\exp(\boldsymbol{v}'_i{}^\top \boldsymbol{v}'/\tau)}{\sum_{j=1}^{K} \exp(\boldsymbol{v}'_j{}^\top \boldsymbol{v}'/\tau)} \tag{4}$$

となる．ここで $\boldsymbol{v}'$ は ID と同様に $||\boldsymbol{v}'|| = 1$ と正規化する．以上より ID と同様に考えて損失関数は

$$\begin{aligned} L_{SD} &= -\sum_{i=1}^{N} \log P(i|\boldsymbol{v}'_i) \\ &= -\sum_{i=1}^{N} \log \frac{\exp(\boldsymbol{v}'_i{}^\top \boldsymbol{v}'/\tau)}{\sum_{j=1}^{K} \exp(\boldsymbol{v}'_j{}^\top \boldsymbol{v}'/\tau)} \end{aligned} \tag{5}$$

となる．

SD は似た波形は同じ特徴マップとして特徴マップの類似度を求める役割を持つ．また SD は前提として同じセグメント内にある波形は似た波形となっていることを期待しており，これによりノイズなどの細かい差異に対して頑健に動作しつつセグメント間の違いを特徴表現として獲得することを期待する．

3.2.4 Autoencoder

AE は Encoder f_{enc} と Decoder f_{dec} からなり，Encoder では加速度センサデータから特徴マップ $\boldsymbol{v} = f_{enc}(\boldsymbol{s})$ に変換を行い，Decoder では特徴マップから加速度センサデータへ復元 $\hat{\boldsymbol{s}} = f_{dec}(\boldsymbol{v})$ を行う．このとき Encoder と Decoder のモデル構造は問わないが，Encoder は VGG のような Convolutional Neural Network を，Decoder はこれを反転したような Deconvolution ベースのモデルを用いることが多い．Encoder と Decoder の両方のパラメータを訓練するために損失関数として入力波形と復元した波形の平均二乗誤差を用いており，ミニバッチ数を M とすると損失関数は以下の式となる．

$$L_{rec} = \frac{1}{M}||\boldsymbol{s} - \hat{\boldsymbol{s}}||^2 = \frac{1}{M}||\boldsymbol{s} - f_{dec}(f_{enc}(\boldsymbol{s}))||^2 \tag{6}$$

AE は Encoder が特徴抽出器として SD とは異なる特徴表現を獲得することを期待する．SD では特徴マップの類似度を測り分類をベースとした特徴表現を Encoder が獲得すると考えられるが，AE を用いることで加速度センサデータを復元するという異なるタスクによる特徴表現を Encoder が獲得できると考える．したがって Encoder が特徴抽出器としてより豊かな特徴表現を得られることを期待する．

3.2.5 モデル全体の訓練手順

以上から提案するモデルの損失関数は FIS の損失関数 L_F と SD の損失関数 L_{SD}，AE で用いる復元誤差 L_{rec} を組み合わせた関数

$$L = \lambda_1 L_F + \lambda_2 L_{SD} + \lambda_3 L_{rec} \tag{7}$$

を訓練に用いる．今回の実験では $\lambda_1 = \lambda_2 = \lambda_3 = 1$ とした．

4. 実験設定

4.1 データセット

提案手法の行動認識に対する有効性を実験により検証する．実験は SD を利用するのに適した行動認識のベンチマークデータセットである HASC を用いて行う．HASC は複数の人により計測された比較的大規模な行動認識のデータセットであり，特徴として様々なデバイスや測定位置により計測されている．複数のデバイスで測定されているためサンプリング周波数などのセンサ特性が混在している．それぞれの計測データに対して「歩行」「走行」「スキップ」「階段上り」「階段下り」「静止」の 6 つの行動ラベルがアノテーションされている．

実験では HASC データセットのうちサンプリング周波数と端末の制限を行った．使用する加速度センサデータは Apple 製の端末によって計測されたサンプリング周波数 100 Hz の加速度センサデータを使用する．被験者はランダムに選択を行い，90 人の計測データをラベルなし大規模データセット D_{unsup} とし，5 人を少量の教師ありデータセット D_{sup}，30 人を検証用データセット D_{test} として各データセットを作成する．

そしてデータの前処理はスライディングウィンドウ方式で加速度センサデータの分割と SD に使用するラベルの作成を行う．加速度センサデータの分割ではウィンドウサイズとストライド幅はともに 512 サンプルとする．SD に使用するラベルの作成は加速度センサデータの分割と同時に行う．

4.2 データ拡張

モデルの訓練に使用するデータ拡張は swapping と flipping を適用する．swapping は加速度センサデータの軸をランダムに入れ替える手法で，例として入力する加速度センサデータの x 軸，y 軸，z 軸を $\boldsymbol{s} = [\boldsymbol{s}_x, \boldsymbol{s}_y, \boldsymbol{s}_z]$ と表現するとき，ランダムに入れ替えた後の加速度センサデータは $\hat{\boldsymbol{s}} = [\boldsymbol{s}_z, \boldsymbol{s}_y, \boldsymbol{s}_x]$ のように処理を行う．flipping は加速度センサデータの符号の反転をランダムに適用する．本実験では適用したときの加速度センサデータの符号はすべて反転させ，このときの加速度センサデータは $\hat{\boldsymbol{s}} = -1\boldsymbol{s}$ となる．

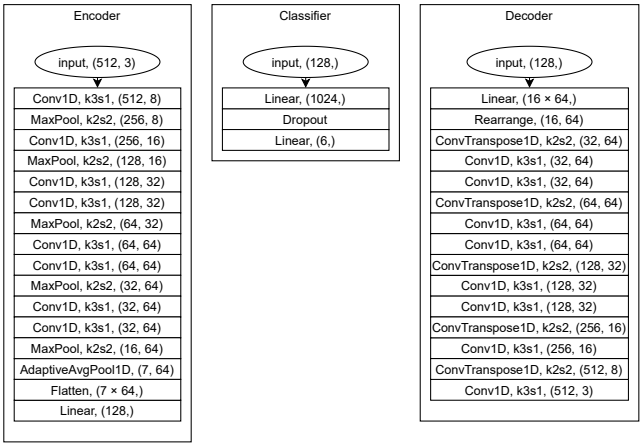

図 4 実験に使用したモデル構造．図中の k3s1 は kernel size が 3，ストライド幅が 1 であることを表している．カッコ内の数値はその層が出力するデータの shape を表している．

Fig. 4 The model architecture used in experiments. “k3s1” means that the kernel size is 3 and stride is 1. The numbers in parentheses denotes the shape of the data output by the layer.

4.3 モデル構造と訓練手法

実験で使用したモデルを**図 4** に示す．図中の k3s1 などは kernel size が 3，ストライド幅が 1 であることを表し，カッコ内の数値はその層が出力するデータの shape を表している．Encoder は VGG11 をベースに加速度センサを扱えるように Conv2D ではなく Conv1D に変更している．Decoder は Deconvolution を用いて Encoder の入力と出力を反転させたような構造とした．また，追加訓練を行うときとベースラインとして使用するモデル構造は Encoder の末尾に Classifier を付けた構造をしており，Classifier は 2 層の Linear 層からなるモデルで，過学習を防止するために Dropout を Linear 層の間に挿入する．

使用するモデルは損失関数と最適化関数 Adam を用いて訓練する．大規模データセット D_{unsup} を用いた表現学習の訓練は式 (7) で提案した損失関数 L を用いる．このときの訓練に使用するパラメータとしてエポック数は 1000，バッチサイズは 128，学習率は 0.001 とした．FT は事前訓練済みの Encoder の末尾に Classifier を付け，損失関数として Softmax Cross Entropy を用いて少量のデータセット D_{sup} で訓練を行う．このとき Encoder のパラメータは固定しない．そして訓練に使用するパラメータはエポック数を 100 とし，ほかのパラメータは事前学習の訓練と同様に設定した．ベースラインとして，事前訓練なしに小規模データセット D_{sup} のみを用いた訓練を行うモデルも実験する．訓練設定はエポック数を 300 として，ほかのパラメータは追加訓練と同様に設定した．

4.4 評価指標

評価指標はクラスタリング精度（Clst 精度）と追加訓練

であるファインチューニングを行ったときの精度（FT 精度）を用い，シード値を固定しプログラム内で連続して 10 試行した際の平均値で議論を行う．Clst 精度は大規模データセット D_{unsup} を用いて訓練した Encoder から得られる特徴マップをクラスタリングすることで測る．クラスタリング手法として k-means 法を用い，クラス数は HASC の行動ラベルに合わせて 6 とした．Clst 精度を計算するために，クラスタリング結果と実際の行動ラベルを比較して精度が最大となるようにクラスタリング結果に行動ラベルを割り当てる．このときの最大の精度が Clst 精度となる．FT 精度は D_{unsup} を用いて追加訓練したモデルの予測と真のラベルを比較することで測る．モデルの予測と真のラベルが一致した箇所をカウントし，全体のデータ数で割ることで精度を算出する．またベースラインモデルの精度は FT 精度と同様に計算する．

加速度センサデータから求めた特徴マップの行動認識における有効性を上述の 2 指標を用いて検証する．Clst 精度が高ければ加速度センサデータに存在する行動の特徴がよく反映された特徴マップであるため行動認識に有用であり，低ければ行動認識に適さない特徴マップだと考えられる．そして，FT 精度が高ければ事前に獲得した特徴表現が行動認識に有効であることを表すと考えられ，少量データセットのみで訓練したベースラインモデルを比較することで事前訓練の必要性を考察する．

5. 実験

5.1 Ablation Study

提案手法の Clst 精度と FT 精度，そして提案手法の組み合わせを個別に検証したそれぞれの精度の結果とベースラインモデルの結果を**表 1** に示す．提案手法は表 1 の最後の行にあり，何も選択されていない最初の行はベースラインモデルを表す．そのほかの行は提案手法の組み合わせを個別に検証した結果を表す．また，Clst 精度と FT 精度の最も高かった結果をそれぞれ太字で表している．Clst 精度を比較すると AE だけの場合が 23.0 % と最も低く，提案手法が 66.2 % と最も高い精度を示した．提案手法と SD を ID に変更した結果を比較すると ID を用いた場合と比べて 13.3 % 精度が向上した．また，AE を用いない場合と比べて 11.5 % 精度が向上した．FT 精度は提案手法が 77.4 %，ID に変更した結果が 76.1 % と 1.3 % 高い精度を示した．またベースラインモデルと比較して提案手法が 3.0 % 高い結果を示した．

Clst 精度の結果から提案手法が最も行動認識に有用な特徴表現を獲得できた．ID を SD に変更した場合の結果に着目すると FIS と ID と比較して精度が向上しており，FIS と ID，AE と比較しても精度が向上していることから似た波形が同じ特徴となるように訓練したほうが行動認識において有用な特徴表現の獲得に貢献できると考えられる．ID

表 1 提案手法の組み合わせによる精度.
Table 1 ablation study.

FIS	ID or SD	AE	Clst 精度	FT 精度	少量のみ
-	-	-	-	-	74.4
-	-	✓	23.0	67.1	-
✓	ID	-	38.0	65.7	-
✓	SD	-	54.7	73.0	-
✓	ID	✓	52.9	76.1	-
✓	SD	✓	**66.2**	**77.4**	-

ではインスタンスごとに分類しようとするため加速度センサデータに含まれるノイズの影響を SD より強く受けたのではないかとも考えられる．そして，AE の有無に関して着目すると AE がある場合はない場合と比べてどれも精度が向上していることから ID や SD などの特徴表現の距離をベースとした表現学習手法だけでなく復元タスクなどの異なるタスクの表現学習手法を組み合わせることはより有用な特徴表現の獲得につながることを示唆すると考える．FT 精度においても Clst 精度と同様な結果の傾向があるため教師なし表現学習手法で獲得した特徴表現は追加訓練を行う上でも有用な特徴表現であると考えられる．また，提案手法はベースラインモデルの結果と比較して FT 精度が高いためラベルなし大規模データセットを用いた事前訓練を行うことで精度が向上することを示した．

5.2 特徴マップの次元数の調査

特徴マップの次元数によって提案手法の精度がどのように変化するのか調査を行った結果を**図 5** に示す．図 5 は縦軸が精度，横軸は Encoder の出力する特徴マップの次元数，エラーバーは標準偏差を表す．凡例の clustering は Clst 精度を表し，fintuning は FT 精度，baseline はベースラインモデルの精度を表している．Clst 精度は次元数が 32 のとき 68.2 % と最も高く，128, 256 のとき 66.2 %, 66.3 % となった．FT 精度は次元数が 16 のとき 73.5 % でそのほかの次元数での精度は約 77 % と大きな変化がなかった．

結果から Clst 精度だけに着目すると次元数 32 が最も高かったが次元数 128, 256 においても比較的高い精度であった．FT 精度は次元数を 32 より大きくしても大きな変化は見られなかったが，次元数が 128 のときに精度が 77.4 % と最も高く，標準偏差が 1.8 と比較的小さい値となった．次元数 16 では Clst 精度が低いため，行動認識に有用な特徴表現が獲得できず結果として追加訓練を行っても高い精度が得られなかったと考えられる．ベースラインモデルではどの次元数においても大きな変化は見られなかった．また Clst 精度と FT 精度を考慮して本実験では次元数 128 を選択した．

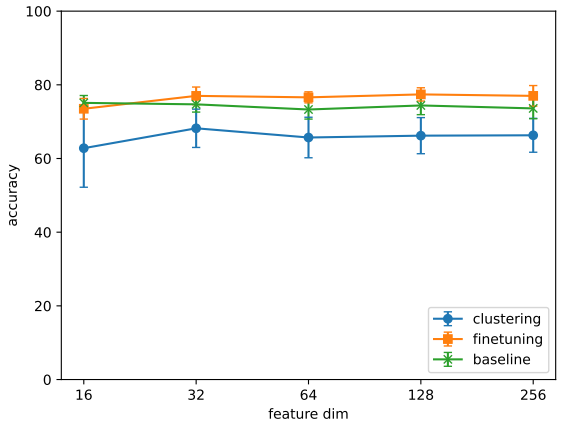

図 5 特徴マップの次元数の変化による精度の変化．横軸は Encoder が出力する次元数で縦軸は精度，エラーバーは標準偏差を表す．

Fig. 5 Accuracies with the number of dimensions of the feature. The horizontal axis shows the number of dimensions, the vertical axis shows the accuracy, and error bar means the standard deviation.

6. おわりに

本研究では加速度センサデータの教師なし表現学習に取り組んだ．画像分野で提案された ID と FIS を用いた手法をベースとして ID からセンサに適した SD に変更し，AE と組み合わせた手法を提案した．行動認識ベンチマークデータセット HASC を用いた実験を行い，提案手法は Clst 精度が 66.2 % と最も高い精度を示した．ファインチューニングを行う実験では分類精度が 77.4 % となり，少量データのみで訓練したベースラインモデルより高い精度を達成した．結果から提案手法が加速度センサデータから行動認識に有用な特徴表現を獲得できることを示した．

今後の課題として，実験では HASC にあるすでに分割された加速度センサデータを使用したが，暗黙的にラベル情報が含まれた可能性があるためシーケンシャルなデータを用いて特徴表現が獲得できることを示すことがあげられる．

謝辞 本研究の一部は，JSPS 科学研究費助成事業若手研究 (19K20420) の助成によるものである．ここに謝意を表す．

参考文献

[1] Zeng, M., Nguyen, L. T., Yu, B., Mengshoel, O. J., Zhu, J., Wu, P. and Zhang, J.: Convolutional Neural Networks for human activity recognition using mobile sensors, *6th International Conference on Mobile Computing, Applications and Services*, pp. 197–205 (online), DOI: 10.4108/icst.mobicase.2014.257786 (2014).

[2] Murad, A. and Pyun, J.-Y.: Deep Recurrent Neural Networks for Human Activity Recognition, *Sensors*, Vol. 17, No. 11 (online), DOI: 10.3390/s17112556 (2017).

[3] Mahmud, T., Sazzad Sayyed, A. Q. M., Fattah, S. A. and Kung, S.-Y.: A Novel Multi-Stage Training Approach for Human Activity Recognition From Multimodal Wearable Sensor Data Using Deep Neural Network, *IEEE Sensors Journal*, Vol. 21, No. 2, pp. 1715–1726 (online), DOI: 10.1109/JSEN.2020.3015781 (2021).

[4] Saeed, A., Ozcelebi, T. and Lukkien, J.: Multi-Task Self-Supervised Learning for Human Activity Detection, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 3, No. 2 (online), DOI: 10.1145/3328932 (2019).

[5] Haresamudram, H., Essa, I. and Plötz, T.: Contrastive Predictive Coding for Human Activity Recognition, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 5, No. 2 (online), DOI: 10.1145/3463506 (2021).

[6] van den Oord, A., Li, Y. and Vinyals, O.: Representation Learning with Contrastive Predictive Coding, *CoRR*, Vol. abs/1807.03748 (online), available from ⟨http://arxiv.org/abs/1807.03748⟩ (2018).

[7] Sheng, T. and Huber, M.: Unsupervised Embedding Learning for Human Activity Recognition Using Wearable Sensor Data (2020).

[8] Wu, Z., Xiong, Y., Yu, S. X. and Lin, D.: Unsupervised Feature Learning via Non-parametric Instance Discrimination, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Los Alamitos, CA, USA, IEEE Computer Society, pp. 3733–3742 (online), DOI: 10.1109/CVPR.2018.00393 (2018).

[9] Gansbeke, W. V., Vandenhende, S., Georgoulis, S., Proesmans, M. and Gool, L. V.: SCAN: Learning To Classify Images Without Labels, *ECCV* (2020).

[10] Tao, Y., Takagi, K. and Nakata, K.: Clustering-friendly Representation Learning via Instance Discrimination and Feature Decorrelation, *International Conference on Learning Representations*, (online), available from ⟨https://openreview.net/forum?id=e12NDM7wkEY⟩ (2021).