［研究助成（A）］

深層生成モデルを用いたセンサデータの超解像と行動認識への応用

Super-Resolution of Sensor data using deep generative model
towards applying activity recognition

2191027

研究代表者　　福井大学　学術研究院工学系部門　　講　師　　長谷川　達　人

［研究の目的］

スマートデバイスやIoT機器が普及した現在，周囲の環境や人間の行動などをコンピュータに認識させるコンテキストアウェアネスに関する研究が盛んに行われている。中でも，スマートフォン等に搭載されているセンサを用いて，ユーザの身体活動の自動認識を行う行動認識は多くの研究がなされている。行動認識が実現できることで，ライフログや，行動に応じたスマートフォンの機能変更，大規模なユーザデータを集めれば，マーケティングや集団行動の解析など，様々な分野への応用が可能となる。

行動認識は，設置型のカメラを用いて画像ベースに行う手法もあるが，本研究では，ユーザが常日頃から持ち歩く機器で計測したセンサデータを用いた行動認識に焦点を当てる。設置型と比べ，時間や場所を問わずデータが取得できる点や，特定のユーザに特化した情報を獲得しやすいというメリットがある。各ユーザが独自にデバイスを所持しなければならない手間がデメリットだが，近年急激に普及したスマートフォンやスマートウォッチを利用することで回避できる。しかし，デバイスの種類，所持方法，装着方法，計測アプリケーション等の様々な計測条件が，ユーザによって，計測日によって異なるという問題点が残る。

センサデータを機械学習で自動分類する手法が一般的な行動認識の原理である[1]。一方，機械学習では一貫性のないデータを用いてモデルを訓練すると推定精度が低下する恐れがあるため，計測環境は統一されていることが望ましい。深層学習では訓練時に大量のデータが必要となるが，計測環境を統一しつつ大量のデータを収集することは容易ではない。すなわち，一貫性の無いデータに対しても高い認識精度を達成できる手法が望まれる。

以上を踏まえ，本研究ではセンサを用いた行動認識において，様々な計測環境に対して頑健な行動認識手法を開発することを目的とする。本稿では特に，サンプリング周波数に焦点をあてる。サンプリング周波数はデバイスの種類や計測アプリケーションによって変動し行動認識精度に影響を与える要因の一つである。例えばAndroidスマートフォンの場合，APIでサンプリング間隔を変更することができる。機種やスペックによってもサンプリング間隔が多少変動する。ウェアラブルデバイスを併用する場合なども変動する。そこで，本研究ではサンプリング周波数の相違に対して頑健な行動認識手法を開発する。深層学習による敵対的訓練を応用することで，様々なサンプリング周波数で計測されたセンサデータが混在する環境（以降，SF混在環境とする）における，行動認識精度の向上を実現する。

［研究の内容，成果］

1. 主要な貢献

以下 3 点が本研究の主要な貢献である。

・従来研究では，実験内でサンプリング周波数を統一して研究が行われていた。本研究では，SF 混在環境が行動認識精度に与える影響を実験により明らかにした。

・深層学習を用いた行動認識の識別器に対して，サンプリング周波数の弁別器を敵対させる行動認識手法を提案し，定式化を行った上で実装した。深層学習を用いた行動認識モデルに SF 混在環境のデータセットを入力した際，モデルの表現力の高さからサンプリング周波数ごとに特徴表現を獲得する可能性がある。そこで，あえてサンプリング周波数が弁別できないような特徴表現を獲得させることで，SF 混在環境における頑健性を高められると考え，本手法の着想に至った。

・複数のベースラインモデルと提案手法を実装し，SF 混在環境における行動認識精度を比較検証した。結果，提案手法が推定精度を向上させることを明らかにした。最終的に，ダウンサンプリングしたデータを併用した提案手法で最高精度を達成した。

・サンプリング周波数毎にデータ数の偏り方が異なるケースについても検証を行い，どの条件下においても提案手法が有効に働くこと，及びその偏りごとの特徴を明らかにした。

2. 提案手法

2.1 モデル構造

Li らの研究[1]のように，従来の研究で採用されている深層学習を用いた行動認識モデルは，入力データをシンプルにネットワークに通し，行動を分類するモデルである。ネットワークは Multi Layer Perceptron（MLP）ベースのもの，Convolutional Neural Network（CNN）ベースのもの，Recurrent Neural Network（RNN）ベースのもの等様々だが，本稿では CNN ベースで議論を行う。前述の通り関連研究ではこのモデルを特定のサンプリング周波数に限定して訓練を行っていた。従って，SF 混在環境にシンプルに応用する場合，図 1 のようにサンプリング周波数ごとにモデルを訓練することになる。ここで，$\mathbf{X}_{100\,\mathrm{Hz}}$はサンプリング周波数 100 Hz で計測されたセンサデータであり，ネットワークへの入力を意味する。$\mathrm{E}_{100\,\mathrm{Hz}}$は同 100 Hz 用に訓練される特徴抽出器，$\mathbf{z}_{100\,\mathrm{Hz}}$は$\mathrm{E}_{100\,\mathrm{Hz}}$から出力される特徴マップ，$\mathrm{C}_{100\,\mathrm{Hz}}$は行動を分類する分類器，$\hat{y}$は分類器の予測結果である。予測結果$\hat{y}$と真の行動ラベル$y$から，クロスエントロピー誤差$\mathrm{L}_{\mathrm{CE}}$を算出し，ネットワークの損失関数としている。従来手法は，サンプリング周波数ごとにモデルを訓練，管理しなければならない点がデメリットとなる。また，サンプリング周波数ごとにモデルを訓練することから訓練データの量が限定的となり，深層学習モデルを訓練しきれない可能性もある。

提案手法のモデル構造を図 2 に示す。特徴抽出器 E は全てのサンプリング周波数の計測データを一手に入力として受け入れる。この時，入力長を統一するため，$\mathrm{X}_{50\,\mathrm{Hz}}$等のデータは全て 100 Hz になるように線形補間を用いてアッ

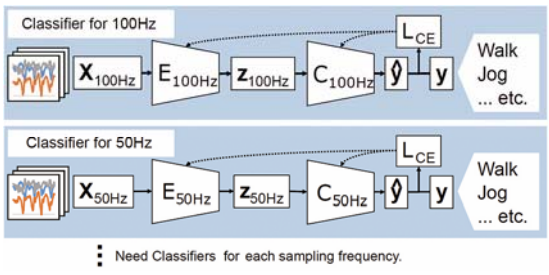

図 1　従来手法：CNN を用いた一般的な行動認識

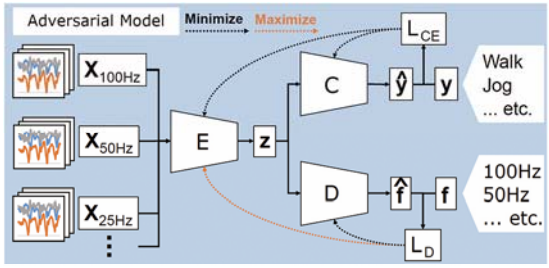

図 2　提案手法：サンプリング周波数に頑健な行動認識

プサンプリング処理を行っている。すなわち，入力データの時間長，サンプル長は前処理として揃えている。次に，図２の上部では従来手法と同様に行動認識を行い，L_{CE}を最小化するようにＥとＣを訓練する。図２の下部では，$\mathbf{z}$をＤにも同時に入力し，$\mathbf{z}$からサンプリング周波数fを予測する。これはドメイン適応手法DANN[2]におけるドメイン検出部と同様の働きをする。本モデルでは入力長は線形補間により統一されているが，$X_{100\,Hz}$と線形補間済みの$X_{25\,Hz}$では波形の滑らかさに差がある。Ｄはこの差を検出することでサンプリング周波数を推定する。予測結果$\hat{f}$と真のサンプリング周波数fから，Discriminator 誤差L_Dを算出し，ネットワークの損失関数としている。本提案手法のメリットは，サンプリング周波数ごとに分割せず唯一のモデルとして訓練ができる点，サンプリング周波数が混在するデータ全てを利用することによる精度向上が見込める点，そして敵対的訓練によりサンプリング周波数の相違に頑健な特徴表現が獲得できる点である。

2.2　定式化

L_{CE}は従来手法でも用いられている，一般的なニューラルネットワークの損失関数であり，下式(1) で定義される。

$$L_{CE}=-\frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M}y_{nm}\log C\left(E(x_n)\right) \qquad (1)$$

ここで，Ｎはミニバッチサイズ，Ｍは行動ラベルのカテゴリ数，x_nは入力$X_{100\,Hz}$のミニバッチのうちｎ番目の入力系列，y_{nm}は同ｎ番目の出力を one-hot 表現した際のｍ番目の値{0, 1}である。Ｃ及びＥは従来手法であれば各周波数に対応したものである必要がある（例：$C_{100\,Hz}$）。

L_Dはサンプリング周波数を弁別する Discriminator から得られる損失関数であり，本研究ではこれを次式(2) で定義する。

$$L_D=-\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K}f_{nk}\log D\left(E(x_n)\right) \qquad (2)$$

ここで，Ｋはサンプリング周波数の種類数である。すなわちL_Dはサンプリング周波数を分類するＤのクロスエントロピー誤差である。DANN[2] では，ソースとターゲットドメインを弁別するためバイナリクロスエントロピー誤差が採用されていた。一方，本研究ではいくつかの限定された種類のサンプリング周波数が混在する環境を想定し，これをカテゴリ分類問題として取り扱うこととした。

以上を踏まえ，提案手法のネットワーク全体の損失関数Ｌを下式(3) で定義する。

$$L=L_{CE}-\lambda L_D \qquad (3)$$

今回，λは１としている。

2.3　最適化

提案手法は，特徴抽出器Ｅ，行動分類器Ｃ，サンプリング周波数弁別器Ｄの３つのモデルで構成される。それぞれのネットワークのパラメータを$\theta_E, \theta_C, \theta_D$とすると，探索対象の各パラメータ$\hat{\theta}_E\,\hat{\theta}_C, \hat{\theta}_D$は下式(4, 5) で算出される。

$$\left(\hat{\theta}_E, \hat{\theta}_C\right)=\mathrm{argmin}_{\theta_E,\,\theta_C}\,L\left(\theta_E, \theta_C, \hat{\theta}_D\right) \qquad (4)$$

$$\hat{\theta}_D=\mathrm{argmax}_{\theta_D}\,L\left(\hat{\theta}_E, \hat{\theta}_C, \theta_D\right) \qquad (5)$$

本研究では，上記の最小化，最大化問題を交互最適化により実装する。式(4) では，θ_Dを固定した状態で算出されたLを最小化するようにθ_E, θ_Cを探索する。これは，L_{CE}の最小化と，L_Dの最大化を行うようなパラメータを探索しており，行動認識精度を上げるようにＥ，Ｃを訓練し，サンプリング周波数の弁別精度を下げるようにＥを訓練することと同義である。式(5) では，同様にθ_E, θ_Cを固定した状態で算出されたLを最大化するようにθ_Dを探索する。この時，パラメータが固定されているためL_{CE}は変動しないことから，Lの最大化はL_Dの最

小化を意味している。したがって，サンプリング周波数の弁別精度を上げるようにDを訓練することと同義である。ミニバッチ単位でこれらを交互に最適化する。

3. 評価実験

3.1 実験環境

スマートフォンの加速度センサを用いて人間の基本行動認識を行うHASCデータセット[3]を用いる。HASCは人間行動理解のための装着型センサによる大規模データベースの構築を目的とした非営利任意団体が提供する行動認識データセットである。基本行動6種類（停止，歩行，走行，スキップ，階段上り，階段下り）のラベルがついた加速度，ジャイロ等のセンサデータがコーパスとして提供されている。

2011から2013年までのコーパスのBasic Activityよりサンプリング周波数が100 Hzのデータを抽出し，加速度センサの生データのみを用いることとした。前処理として，各計測ファイルから前後5秒を除去し，フレームサイズ256サンプル，ストライド256サンプルで時系列分割を行う。計測開始から端末の格納動作等の影響を取り除くため前後5秒でトリミングしている。計測機種や性別等のメタ情報は用いない。トリミング後に1フレーム以上データが取得できた176名のデータを採用した。

行動認識研究では，評価対象者自身のデータを訓練時に使用することで推定精度が向上することが知られている。したがって本研究では，訓練と検証用のデータセットを人単位で分けるHold-out法により分割する。176名のデータセットからランダムに100名を抽出し訓練用データセットD_{train}とし，同様に別途ランダム抽出した50名を検証用データセットD_{test}とする。D_{test}はダウンサンプリングによって100 Hzのものから50 Hz，25 Hz，12.5 Hz，6.25 Hzと複数種類準備し，それぞれで精度検証を行う。

訓練用データセットはD_{train}からダウンサンプリングによりSF混在環境を疑似的に再現する。はじめに，D_{train}を特定の人数で分割する（A，B，C，D，E名）。A名のデータはサンプリング周波数100 Hzで計測されたものとし，B名のデータはサンプリング周波数50 Hzで計測されたものとし，ダウンサンプリングにより生成する。B名の100 HzデータはD_{train}自体には含まれているが，100 Hzの状態では実験に使用しない。同様にC名は25 Hzで，D名は12.5 Hzで，E名は6.25 Hzで計測されたデータとしてダウンサンプリングにより生成する（D'_{train}とする）。

D'_{train}は更にダウンサンプリングを行うことにより，自身より低サンプリングレートのデータセットを擬似的に再現することが可能となる。A名の100 Hzデータからは，50 Hzや25 Hz等を再現できる。これを全パターン実施したデータを$D'_{full\text{-}train}$とする。

3.2 ベースライン

ベースラインとして訓練用データの扱い方を工夫し以下の5種類を定義する。各名称のカッコ内の数字は，訓練が必要なモデル数を意味している。

Ref（5） 従来手法と同様，検証用データのサンプリング周波数と一致するデータのみをD'_{train}から使用しモデルを訓練する手法である。サンプリング周波数毎にモデルを使い分ける。

Mixin（1） D'_{train}全てを用いてモデルを訓練する手法である。シンプルに適応用データを混在させるMixinで高精度を達成することが示されている[4]ことから，ベースラインとして採用した。

DS-Mixin（5） Down Sampling Mixinを意味する。基本はMixinだがダウンサンプリングにより検証用データのサンプリング周波数と一致させられる際には一致させたデータを訓練に用いて，5つのモデルを使い分ける手法である。

F-Mixin（1） Full Mixinを意味する。$D'_{full\text{-}train}$

全て Mixin する手法である。シンプルにデータ拡張による精度向上が見込める観点から採用した。

Multi（1）行動認識とサンプリング周波数弁別を同時に実施するマルチタスク学習を行う手法である。サンプリング周波数の学習をモデル内で行う手法のベースラインとして採用した。

これに対して，敵対的訓練を用いた提案手法も，以下の 2 種類定義する。

Adv（1）D'_{train}を用いてシンプルに提案モデルを訓練する手法である。

F-Adv（1）Full Adversarial を意味する。F-Mixin のように$D'_{full\text{-}train}$を全て用いて提案モデルを訓練する手法である。

3.3　実験結果

20 試行実施した平均推定精度を表 1 に示す。1 行目は検証対象のD_{test}のサンプリング周波数である。各検証に対して，最大精度を達成できた手法を太字かつ下線で，2 番目に精度が高かった手法を下線で示している。

表 1 より，検証用データが 100 Hz から 12.5 Hz は提案手法の **F-Adv** が最高精度，6.25 Hz は **DS-Mixin** が最高精度を記録した。更に，検証用データが 6.25 Hz であっても，**F-Adv** は **DS-Mixin** とほぼ同等の精度を記録している。2 番目に精度が高かった手法に着目すると，検証用データのサンプリング周波数が高いときは **Adv** が，低いときは **DS-Mixin** や **F-Mixin** が高精度を達成した。考えてみると当然ではあるが，検証用データが 6.25 Hz の際にはダウンサンプリングにより 100 名分の訓練データが揃った状態となるため，**DS-Mixin** が十分な特徴表現を獲得し最高精度を達成できた。一方で，センサデータに対する超解像技術が確立されない限り，基本的に高サンプリング周波数側にデータ拡張を行うことができないため，**DS-Mixin** や **F-Mixin** は **Mixin** と変わらない推定精度となった。

表 1　20 試行を行った際の各手法の平均推定精度［%］

Method＼Test data	100Hz	50Hz	25Hz	12Hz	6Hz	Avg
Ref	74.9%	75.5%	73.6%	75.4%	68.9%	73.6%
Mixin	85.6%	85.9%	85.9%	85.1%	75.9%	83.7%
F-Mixin	86.2%	86.4%	86.5%	86.3%	82.7%	85.6%
DS-Mixin	85.5%	85.8%	86.1%	86.3%	**83.0%**	85.3%
Multi	83.6%	84.0%	84.1%	82.0%	72.1%	81.2%
Adv	86.3%	86.5%	86.4%	85.4%	76.0%	84.1%
F-Adv	**86.8%**	**86.9%**	**87.0%**	**86.6%**	82.5%	86.0%

［今後の研究の方向，課題］

本研究では，センシングによる行動認識を対象に，計測データのサンプリング周波数が混在する環境に頑健な行動認識手法の開発を行った。評価実験を行った結果，提案手法の有効性を示したが，スマートフォンの行動認識データセットにおける有効性検証のみである点，各サンプリング周波数はダウンサンプリングにより擬似再現している点等，いくつかのリミテーションがある。今後はこれらの点に関する検証実験を行う予定である。

なお，本研究の詳細は成果発表［A］にて論じている。また，当初計画ではセンサデータに対する超解像によりデータを鮮明化し，行動認識精度向上を図る予定であった。こちらの詳細については成果発表［B］にて論じている。

［参 考 文 献］

［1］ Frédéric Li, et al. : Comparison of Feature Learning Methods for Human Activity Recognition Using Wearable Sensors, Sensors, Vol. 18, No. 679, pp. 1－22（2018）.

［2］ Ganin, Y., et al. : Domain-Adversarial Training of Neural Networks, The Journal of Machine Learning Research, Vol. 17, No. 1, pp. 2096－2030（2016）.

［3］ Kawaguchi, N., et al. : HASC Challenge : Gathering Large Scale Human Activity Corpus for the Real-World Activity Understandings, In Proceedings of the Augmented Human International Conference（2011）.

［4］ 岩澤有祐，他：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，Vol. 32, No. 3, pp. 1－11（2017）.

［成果の発表，論文等］

［A］ 長谷川達人，木村洋文：敵対的訓練を用いたサンプリング周波数の相違に頑健な行動認識，DICOMO2020，オンライン開催，2020.（優秀プレゼンテーション賞 受賞）

［B］ 武仲紘輝，長谷川達人：U-Net モデルを用いた加速度センサデータの超解像，FIT2020，オンライン開催，2020.