

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

An Application of Representation Learning to
Human Behavior Sensing of Wheelchair Users with an Accelerometer

岩澤 有祐^{*1} 矢入 郁子^{*2} 松尾 豊^{*1}

Yusuke Iwasawa Ikuko Eguchi Yairi Yutaka Matsuo

^{*1}東京大学工学系研究科技術経営戦略学専攻

Graduate School of Technology Management for Innovation, The University of Tokyo

^{*2}上智大学理工学研究科理工学専攻

Graduate School of Science and Engineering, Sophia University

Recent expansion of intelligent gadgets, such as smart phones and wristwatch shaped vital sensors, boost a good relationship between human behavior sensing in daily lives and useful applications in ubiquitous computing. Representation of the sensing data of human behavior has been depended on handmade feature designs based on domain knowledges or careful observation of data. This paper reports effectivity of representation learning to human behavior sensing with accelerometers by driving logs of nine wheelchair users.

1. はじめに

近年スマートフォンやリストバンドタイプのセンサの普及により、人間行動センシング技術およびその応用が大きな研究テーマの一つとなっている。我々は、人間行動センシング技術を応用した車いすセンシングと車いす行動分類モデルの構築によるアクセシビリティ情報の抽出に関する研究を行ってきた [Iwasawa 12]。一般に、行動分類モデルを作成する場合新しいユーザに頑健なモデルをどのように作成するかが大きな問題の1つとして知られている [Lara 13]。例えば、歩くという行動1つとってみてもその歩幅や足を上げるタイミングなどは人によって異なると考えられる。一方で、行動データ1人1人個別にモデルを作成することは教師データ作成コストの観点から現実的ではない。そのため、モデルを作成する際に新しいユーザにも頑健なモデルを作成することが重要であると考えられる。

新しいユーザに頑健なモデルを作成するための一つのアプローチとして、複数のユーザのデータに対してよい予測性を持つような頑健な特徴表現を利用する方法が考えられる。従来、行動センシングデータから知識を獲得する際には、時間領域での統計量やFFTの係数など、人間による経験的な特徴量の設計が行われてきた [Lara 13, Bulling 14]。表現学習手法を利用し、複数人のユーザに共通の特徴表現を自動で獲得することができれば、従来の経験的な特徴表現の設計以上により特徴表現を獲得できる蓋然性がある。本稿では、車いす行動を計測した加速度データから段差を推定する問題に表現学習手法の1種である Convolutional Neural Network (CNN) を適用し、その有効性を報告する。分類精度による定量的な評価を通して、獲得した特徴表現が行動認識研究でよく利用される経験的な特徴表現と比較して新しいユーザに対して頑健な表現となっていることを検証する。データセットとしては、9名の同一経路での行動を1人あたり約50分から60分を iPod touch に搭載された3軸加速度センサで計測したものを利用した。サンプル数は9名合計で7016サンプルであり、そのうち段差サンプルは650サンプルとインバランスなデータであった。

本論文の貢献は次の2点に集約される。1点目は、CNNを利用してインバランスな車いす加速度データから特徴表現を抽出するアプローチを提案した点である。2点目は、実データを利用して、提案アプローチが新しいユーザに頑健な特徴表現を獲得するために有効であることを示した点である。2章では車いすセンシングデータ解析を行動データ解析の文脈から議論し、3章で本稿で提案するCNNを利用した車いす表現学習アプローチについて述べる。4章で利用したデータセットの取得実験、およびに提案アプローチの評価実験について述べる。最後に5章で結論、将来課題を述べ本論文の結びとする。

2. 車いす加速度センサデータの特徴表現

2.1 車いすセンシングによる路面状況推定の位置づけ

本稿では、車いすセンシングを利用したアクセシビリティ情報抽出システムの構築に向けて、加速度センサによって計測した車いす走行行動データを利用して表現学習の有効性を検証する。図1に車いすセンシングを利用したアクセシビリティ情報推定のアーキテクチャを示す。本システムの肝は、車いす行動を計測した加速度データから段差や坂、ユーザの疲労状態などを高精度に推定するモデルの構築である。本問題は、行動を計測した時系列データから行動分類モデルを作成する問題として捉えることができる。ウェアラブルセンサによって計測したデータを使った行動認識モデルの構築は、コンテキストウェアナなサービスや日々の健康管理など様々なアプリケーションへの応用が期待されており重要な研究テーマの1つである。

センサによって計測したデータの行動認識モデルを作成する場合の問題の1つとして、同一のラベルに対する計測データのばらつき (Intraclass Variability) が大きくなりがちであることが挙げられる。特にユーザに依存しないモデルを構築しようとする場合、ユーザごとの行動の仕方の違いが問題となる [Lara 13]。車いすデータの場合、例えばあるユーザは段差の前に速度を緩める、といった知識を利用すればあるユーザに対しては分類精度向上が期待されるが、一方で同様の行動をとらないユーザにとっては分類精度低下を引き起こす可能性がある。また、多くの行動データの場合すべてのユーザに対して教師データを獲得することはコストの観点から現実的ではない。本稿では限られたユーザに対して教師データが付与されている

連絡先: 岩澤有祐, 東京大学工学系研究科技術経営戦略学専攻, 〒113-8656 東京都文京区本郷 7-3-1, iwasawa@weblab.t.u-tokyo.ac.jp

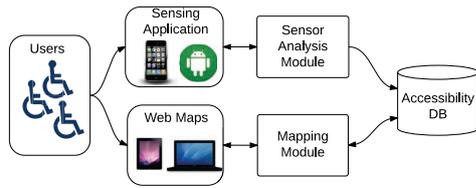


図 1: 車いすセンシングによる路面状況可視化

状況で、教師データの無いユーザに頑健な特徴を獲得するための基礎研究として位置づけられる。

2.2 行動センシング加速度センシングデータの特徴表現

2.2.1 経験的な特徴表現の設計

ほとんどの場合行動センシングデータの特徴表現としてドメインに依存した特徴量の設定が行われてきた [Lara 13, Bulling 14]。よく利用される特徴量の設計を大別すると、1) 時間領域での統計量、2) 周波数領域の特徴量、3) その他の3つの方法がある。Lara らのサーベイ論文によると、時間領域の特徴量としては、平均値、標準偏差値、相関係数などの統計量、ゼロクロッシングや階差などが利用されている [Lara 13]。また、周波数領域では FFT 係数や様々なマザーウェーブレット関数を利用した Wavelet の係数を利用したものなどが代表的な手法として利用されている。その他の方法としては、SAX を利用した方法や、Hammerla らが提案し複数の行動認識タスクでその有効性が確認されている Empirical Cumulative Distribution Function (ECDF) と呼ばれる方法がある [Hammerla 13]。

2.2.2 表現学習

近年いくつかの研究が行動認識タスクへの表現学習手法を適用しその有効性を検討している。表現学習を行動データに適用した際の利点の1つは、データから直接特徴表現を獲得することで経験的に設計するだけでは捉えることができなかった特徴を予測に利用することができる点である。Plotz らは、行動認識のために必要な表現学習方法として PCA と Restricted Boltzman Machine (RBM) を利用したアーキテクチャを提案し、4つの行動認識タスクに対してヒューリスティックに設定した既存の手法と比較して表現学習を用いた手法が有効であることを示している [Plötz 11]。同様に、Vollmer らは Sparse Coding を利用した方法の有効性を示している [Vollmer 13]。Zeng らは、1) 局所的な特徴量が捉えられる、2) 大きさへの不変性を獲得することができる、といった特徴を持つ CNN を利用したアプローチを提案し、RBM や Sparse Coding を利用した方法よりも高い認識精度を達成できることを示している [Zeng 14]。表現学習を行動データに適用した際のもう1つの利点は、ユーザの違いに対して頑健な特徴表現を獲得できる可能性が有る点である。経験的な特徴量の設計方法では、特徴変換を行う際に他のユーザのデータとの関連性は一般に考慮されない (図 2-a)。一方で、表現学習を利用する場合には複数人のユーザの教師データから共通する特徴表現を獲得することによって新しいユーザに対しても頑健な特徴量を獲得することが期待される (図 2-b)。

本稿で CNN を利用したアプローチを車いす加速度データに適用し、特徴表現の新規ユーザに対する頑健性の観点から評価する。本論文は、次の2つの点で既存の研究とは異なっている。1点目はデータの違いである。既存の研究が家の中のような特定の狭い環境の中でとられた健常者の行動データを利用しているのに対して、本研究ではコントロールされていない外部環境で計測したデータへの有効性を確認する。Deep Learning

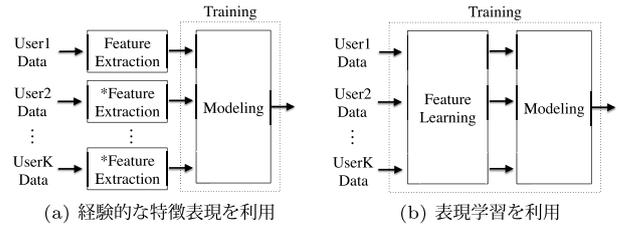


図 2: 経験的な特徴表現の設計と、表現学習を利用した特徴表現の獲得の新規ユーザに対するロバスト性の比較

の応用研究としての観点から見ると、応用範囲を広げるといって意義がある。行動認識研究の観点から見ると、実験環境は認識精度に大きな影響を与えることが報告されており、異なる実験環境での認識精度を評価することは重要である。2点目は、表現学習の新規ユーザに対する頑健性を調査した点である。新規ユーザに対する頑健性は我々のアプリケーションのみならず、行動データを利用するアプリケーションに共通して重要な性質であり検討に値すると考えられる。

3. CNNによる車いす行動センシング加速度データの表現獲得

図3に今回有効性を評価した CNN を利用した表現学習アーキテクチャを示す。提案ネットワークは 1) 入力層、2) 1組以上の畳み込み層とプーリング層の組み合わせ、3) 全結合層、4) 出力層から構成される。入力層には、スライディングウィンドウによって切りだされた生データが入力される。入力されたデータは、畳み込み層とプーリング層、全結合の隠れ層によって特徴表現に変換される。各畳み込み層は入力側の特徴マップ y^{n-1} を n 層目の畳み込みカーネル W^n によって畳み込み新たに $m^n - 1$ 個の特徴マップを次の式に従って出力する。ここで、関数 f は活性化関数、 $m^n - 1$ は $n-1$ 層目の特徴マップ数、 W_{ij} は $n-1$ 層目の特徴マップ i から n 層目の特徴マップ j への重みを表している。

$$y_j^n = f \left(\sum_{i=1}^{m^{n-1}} y_i^{n-1} * W_{ij}^n + b_j^n \right)$$

畳み込み層で得られた出力は、長さ p^{n+1} の Max Pooling で出力した。CNN を利用した他の研究にならい、 $N-1$ 層目をシグモイド関数による全結合層、 N 層目をソフトマックス関数による出力層とした [Simard 03]。 $N-1$ 層目の全結合層は $N-2$ 層目で出力された複数のチャンネルの特徴を 1次元の特徴ベクトルに変換する役割がある。

ネットワークの学習は、バッチサイズ 50 の Mini-Batch SGD によって行った。ただし、一般的に用いられるようにすべてのバッチを順番に利用する方法だとインバランス性の問題から学習結果が多数派のクラスに偏ってしまったため、各バッチで段差クラスが 40%、その他のクラスが 60%となるようにランダムにサンプルを選択した。作成したモデルが特定のユーザにしか使えない状態は、過学習の1種であると考えられる。過学習を避けるためのテクニックの1つである Dropout を利用した [Hinton 12]。各エポックで利用しないユニットは隠層ランダムに 50%とした。また、同様の理由から Weight Decay を利用し過学習を低減した。

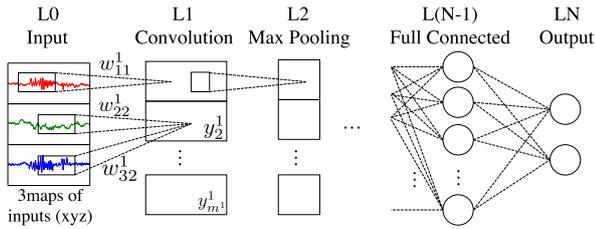


図 3: Convolutional Neural Network を利用した車いすセンシング加速度データの表現学習アーキテクチャ

4. ヒューリスティックな特徴表現との比較

4.1 データセット

車いす行動認識における表現学習の有効性を検証するために、外部環境での移動時の車いす行動を加速度センサで計測した。図 4 に利用したセンシングシステムの外観を示す。iPod touch を車いすの座席下の骨組み部分に取り付け、ゴムバンドで固定した。加速度センサのサンプリング周波数は 50Hz とした。また、椅子の後部には位置情報の計測のための準天頂衛星の受信機をリュックに入れて設置している。加速度センサ、準天頂衛星受信機の他に、ビデオカメラにより実験の様子を記録し、段差を乗り越えているかどうかのラベルとした。実験参加者は合計 9 名であり、男性 7 名女性 2 名、手動車いす 6 名電動車いす 3 名、年齢は 20 代から 60 代と幅広い属性を持つ参加者である。実験では被験者 9 名それぞれに 1 週辺り約 1.5km の経路 1 を 3 週してもらい、移動時の行動を計測した。計測時間は、経路 1 が 1 人あたりおよそ 1 時間 (合計 9 時間)、経路 2 が 1 人あたり 20 分 (合計 180 分) である。できるだけ実際に利用可能なものに近いデータを計測するために、走行者への指示はルート案内にとどめ、スピードやその他の行動などの制限は設けずに計測した。



図 4: Sensing System

4.2 評価方法

表現学習の本問題における有効性の検証の最初の一步として、本稿では段差抽出における精度の比較実験を行った。評価データセットは、既存の行動認識研究にならい [Bulling 14]、計測 3 軸加速度データからスライディングウィンドウ方式により一定時間幅ごとのデータを切り出したものを利用した。スライディングウィンドウの幅として 100, 200, 400 サンプル、重なり率 0.0, 0.25, 0.5 をそれぞれ検討し、比較対象として用いたヒューリスティックな方法で最も精度が高かった 400 サンプル (約 8 秒)、0.5 をすべての特徴量でのパラメタとして利用した。スライディングウィンドウ方式による切り出しの結果得られたサンプル数は合計で 7016 サンプル、うち段差が 650 サンプルとインバランスなデータである。

ベースライン手法として、1) 時間領域の特徴量、2) 周波数領域の特徴量、3) ECDF、の 3 つを利用した。時間領域の特

徴量としては、1) 生データをそのまま利用する方法と、2) 経験的に良い精度を出す方法の 2 つの方法を利用した。経験的な特徴量は xyz の各チャネルの平均値、標準偏差値、最大値、最小値、階差、相関係数、エネルギー、エントロピーなどを予備実験で検討した結果から、xyz の各チャネル及び x の階差、y の階差、z の階差の 6 チャネルそれぞれの平均値、標準偏差値、最大値、最小値の合計 24 の素性を特徴量として計算したものを利用した。周波数領域の特徴量としては、x, y, z の各チャネルの FFT 係数を計算し、各々を結合したものを利用した。

評価方法には、行動認識において新しいユーザに対するモデルの頑健性を検証するためによく利用される方法である Leave-One-Subject 検定を利用した。図 5 に示すように、Leave-One-Subject 検定とは、N 人のデータセットのうち、N-1 人のデータを利用して作ったモデルを残り 1 人のデータを使って精度評価する、ということを N 回繰り返す方法である。この方法では、ユーザ 1 のデータを全く使わずに学習したモデルを使ってユーザ 1 の分類精度を評価するため、新しいユーザに対するモデルの頑健性を評価することができる。今回は、各回のテストデータ i に対する分類精度 E_i をテストデータ数 N_i で重みづけし、その平均値を最終的な評価指標とした。

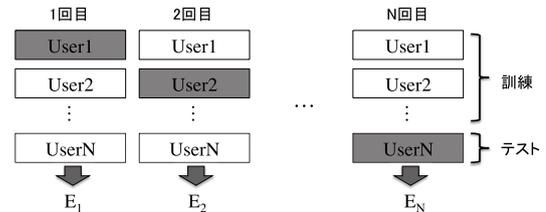


図 5: Leave One Subject Out 検定

4.3 結果

図 6 は、各特徴量の分類結果を比較したグラフである。CNN ベースの特徴量は、畳み込み層 (長さ 25 のフィルタで 150 枚の特徴マップを生成) とプーリング層 (フィルタ長は 4) をそれぞれ 1 層の合計 5 層の構成とした。結果として、今回提案した CNN ベースのアプローチが最も高い精度であった。特に、ヒューリスティックに設定した時間領域での特徴量を除くと、0.2 以上の改善が見られた。ヒューリスティックと比較すると、その差は 0.04 であり大きな差とは言えないが、少なくとも本稿の様な単純な構成でも経験的に設計した以上の分類精度を出せることを示していると言える。

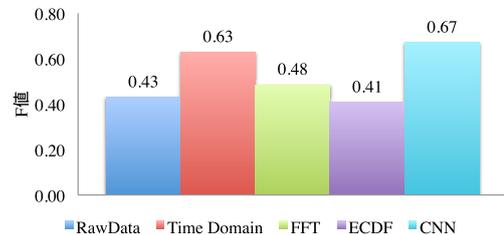


図 6: F 値比較結果

表 1 に提案ネットワークのパラメタを変更した際の訓練誤差と汎化誤差を示す。ネットワーク構成としては、畳み込み層を 1 から 3 層、Kernel サイズを 25 と 45、フィルタ数としては 25, 50, 100, 150, 200 を検討した。結果から、次の 3 つが示された。1) 畳み込み層を 1 層、Kernel サイズを 25、フィルタ数を 150 とする場合に最も汎化性能が良い。ただし、い

ずれの場合でもユーザによって分類精度にばらつきがあることが示されている。2) また、Dropout を利用しない場合だとその他の構成を同様にした場合と比較して 0.15 ポイント精度が低下している。一方で、訓練データでの性能では Dropout を利用しない場合が最も高く、Dropout による過学習の抑制が本問題に重要であることが示唆されている。3) 単純に畳み込み層の数を増やすだけでは分類精度が向上しない。特に訓練データでの評価自体の精度自体が高くなく、学習がうまく行われていない可能性がある。

表 1: CNN のパラメータを変更した際の訓練誤差と汎化誤差

Conv Layers (Total)	Kernel サイズ	フィルタ数	Hidden	Dropout	訓練F値	汎化F値
1(6)	25	25	1024	TRUE	0.77 (0.58-0.83)	0.65 (0.50-0.81)
	25	50	1024	TRUE	0.84 (0.76-0.89)	0.63 (0.38-0.84)
	25	100	1024	TRUE	0.89 (0.67-0.96)	0.66 (0.51-0.87)
	25	150	1024	TRUE	0.93 (0.78-0.97)	0.67 (0.51-0.87)
	25	200	1024	TRUE	0.92 (0.63-0.99)	0.66 (0.40-0.88)
	45	150	1024	TRUE	0.90 (0.75-0.96)	0.66 (0.47-0.84)
	25	150	2048	TRUE	0.92 (0.60-0.98)	0.67 (0.44-0.87)
	25	150	1024	FALSE	0.95 (0.48-1.00)	0.52 (0.08-0.81)
2(8)	25-15	150-150	1024	TRUE	0.80 (0.54-0.91)	0.64 (0.28-0.81)
	25-5	150-150	1024	TRUE	0.75 (0.49-0.91)	0.63 (0.17-0.84)
	25-15	150-250	1024	TRUE	0.73 (0.40-0.92)	0.63 (0.17-0.83)
3(10)	25-15-5	100-150-250	1024	TRUE	0.66 (0.27-0.75)	0.62 (0.05-0.76)
	25-15-5	100-150-250	2048	TRUE	0.70 (0.25-0.80)	0.65 (0.13-0.78)

4.4 考察

4.2 節の結果から、CNN を利用した表現学習を利用した手法が経験的な特徴設計と比較し車いすデータから段差抽出を行う問題において有効であることが示された。本問題は、限られたユーザにのみ教師データが付いている行動センシングデータから、教師データがないユーザに対してもよい予測性能を持つモデルを作成する問題として捉えることができ、提案手法が教師データが入手しにくい一般の行動認識問題についても有効であることが期待される。一方で、本稿では段差という限られた行動に対してのみの有効性の検討にとどまっており、提案したモデルが段差の特徴をよく捉えるモデルであった可能性もある。今後は、同様の性質を持つ幅広いデータに対する有効性の検討が必要である。

また、今回の検証では表現学習手法として CNN を用いたが、本検討がすべての表現学習手法に当てはまるかどうかは検討が必要である。実際、最も単純な表現学習の形の 1 つである主成分分析を用いた方法では期待した精度が得られないことを確認している。よりよい手法の提案に向けて、PCA や ICA などの線形な表現学習のほか、SdA など Deep Learning の他のアーキテクチャでも同様に検討が重要である。

5. 結論と今後の展望

本研究では Convolutional Neural Network による表現学習を車いす加速度データセットに適用し、CNN を使って得られる特徴表現が新しいユーザに対しても頑健な特徴表現の獲得につながっていることを定量的に評価した。結果として、1 層の畳み込みレイヤーによるネットワークでヒューリスティックに設計した特徴量と同程度以上の精度を出せることを確認した。

一方で、1) ユーザごとの分類精度のばらつき、2) 層の数を増やした時の学習精度の低さ、などの問題があった。今後はより多くのデータセットの収集と解析、獲得された特徴表現の定性的評価を通じて、より高精度に複数ユーザに頑健な特徴表現を獲得する手法の提案を行う予定である。

謝辞

走行実験に参加してくださった多くの被験者の方々、実験に協力してくださった方々に感謝いたします。本研究は日本学術振興会特別研究員奨励費の助成によって行われました。

参考文献

- [Bulling 14] Bulling, A., Blanke, U., and Schiele, B.: A tutorial on human activity recognition using body-worn inertial sensors, *ACM Computing Surveys (CSUR)*, Vol. 46, No. 3, p. 33 (2014)
- [Hammerla 13] Hammerla, N. Y., Kirrkham, R., Andras, P., and Ploetz, T.: On preserving statistical characteristics of accelerometry data using their empirical cumulative distribution, in *Proceedings of the 2013 International Symposium on Wearable Computers*, pp. 65–68 ACM (2013)
- [Hinton 12] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R.: Improving neural networks by preventing co-adaptation of feature detectors, *arXiv preprint arXiv:1207.0580* (2012)
- [Iwasawa 12] Iwasawa, Y. and Yairi, I. E.: Life-logging of wheelchair driving on web maps for visualizing potential accidents and incidents, in *PRICAI 2012: Trends in Artificial Intelligence*, pp. 157–169, Springer (2012)
- [Lara 13] Lara, O. D. and Labrador, M. A.: A survey on human activity recognition using wearable sensors, *Communications Surveys & Tutorials, IEEE*, Vol. 15, No. 3, pp. 1192–1209 (2013)
- [Plötz 11] Plötz, T., Hammerla, N. Y., and Olivier, P.: Feature learning for activity recognition in ubiquitous computing, in *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, Vol. 22, p. 1729 (2011)
- [Simard 03] Simard, P. Y., Steinkraus, D., and Platt, J. C.: Best practices for convolutional neural networks applied to visual document analysis, in *2013 12th International Conference on Document Analysis and Recognition*, Vol. 2, pp. 958–958 IEEE Computer Society (2003)
- [Vollmer 13] Vollmer, C., Gross, H.-M., and Eggert, J. P.: Learning features for activity recognition with shift-invariant sparse coding, in *Artificial Neural Networks and Machine Learning-ICANN 2013*, pp. 367–374, Springer (2013)
- [Zeng 14] 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 (2014)