使用智慧手機傳感器進行人類活動識別

Shih Hwa Lai

November 28, 2019

摘要

基於可穿戴傳感器的人類身體活動識別具有與我們的日常生活相關的應用，例如醫療保健，智慧助理，社交網路等。 如何以較低的計算成本實現較高的識別精度是環境感知計算中的重要問題。在這份報告中，我們運用了兩大類機器學習方法，分為淺層和深度學習，來分別處理來自同一個數據集當中，經過特徵工程處理的資料和經過簡單前置處理過的原生訊號。從這兩種不同方法的比較當中分析其優點及缺點。並在結論中討論何種方法較符合未來大數據及「環境感知」計算的趨勢。

目錄

1. 引言 *……………………………. p. 2*
2. 數據集描述 *……………………. p. 2*
3. 數據視覺圖 *……………………. p. 4*
4. 方法 *……………………………. P. 6*
5. 實驗 *……………………………. p.9*
6. 結論 *……………………………. p.12*
7. 參考文獻 *………………………. p.13*

關鍵字：Human Activity Recognition, Time Series Classification, Machine Learning, Deep Learning, Smartphone Sensor, Accelerometer, Gyroscope, Signal Processing

# I. 引言

人類活動識別是實現“ 環境感知運算” 雄心的重要組成部分。在環境計算的框架中，傳感器和計算機被包圍並嵌入到消費者“的日常生活和環境。這種人類傳感器融合的最早例子之一是智能手機的普及。以前，使用移動傳感器是不切實際且昂貴的，因為它們是少量出售的定制硬件。消費者還認為它們是多餘的和非必需的附件，除了感知用戶的動作之外，幾乎沒有任何價值。但是，隨著智能手機的普及帶來的規模經濟，運動傳感器（加速計，陀螺儀和磁力計）的成本下降了。與用於HAR的其他傳感器（視頻，雷達等）方法相比，它已被承諾成為傳感器數據中最易取得的形式。

但是，傳感器的成本只是問題的一半。傳統上，使用信號處理領域的技術來處理來自眾多廉價傳感器的原始信號。這些被用作特徵工程的方法。只有在應用了這些方法之後，才能有效地使用機器學習模型來從數據中學習模式，需要具有專業領域知識的人員來處理。由於這樣專業人員畢竟在人力及成本上有限，也難應付各種不同環境及個人差異的變化。此外，這些模型只能從模式中學習淺層特徵，因此無法很好地概括未見過的數據。因此，如果沒有這些深度的特徵工程，傳統的機器學習方法就不能有效地從原始信號數據中學習。

在此報告中，我們可以證明，通過高質量的特徵工程處理過的數據，使用傳統的機器學習模型（也稱為淺層學習模型）時也可以有很好的精確度。但是，當處理原始傳感器數據時，它們的性能被深度學習模型所超越。

# II. 數據集描述

1. 數據收集及分類

以下對於該數據集的描述是其創建者所給出的。本數據集包括從由30名志願者，年齡從19到48歲所執行6種不同的活動時從掛在腰部的智能手機所採集的。每個志願者在執行這些活動時都將智能手機放在腰部。下面列出了執行的活動及其相應的標籤（請參見 *Table 1.*）。

| Code | Activity |
| --- | --- |
| **1** | Walking |
| **2** | Walking Upstair |
| **3** | Walking Downstair |
| **4** | Sitting |
| **5** | Standing |
| **6** | Laying |
| ***Table 1*. Label for each activity** | |

B. 數據預處理

根據數據集的創建者，信號以50Hz的採樣率記錄下來，並作為每個維度的時間序列存儲，因此獲得了6個不同的信號（3個來自加速度計，其他3個來自陀螺儀）。

對於這些信號中的每一個，使用中值和三階低通20Hz Butterworth [1]濾波器對噪聲進行濾波，以獲得更精確的結果。第二個0.3hz Butterwoth濾波用於將重力效應與加速度計信號分開[1]。

然後，在2.56秒的固定寬度滑動窗口中對信號進行採樣，它們之間有50％的重疊。

原始信號預處理到此結束。在此階段，將數據集導出到每個文件的3個方向的軸（X，Y和Z）中的3個不同的組。第一組是來自智能手機加速度計X軸的總加速度信號，以標準重力單位“ g”表示。每行顯示一個128元素向量。第二類是通過從總加速度中減去重力而獲得的身體加速度信號。最後，第三組是陀螺儀為每個窗口樣本測量的角速度矢量。單位為 (rad/sec)。

在實驗過程中，文件被加載並組合成一個多維數組，其中包含10299條記錄，每個窗口包含了具有50% 重疊的128個 timestep，每個 timestep 包含9個原始預處理信號（請參見 *Table 2.*）。數據集的這一部分將單獨導出，並在 Section *IV B.* 中使用。

C. 特徵工程描述

通過從三軸信號計算歐氏幅度和時間導數（加加速度da / dt和角加速度dw / dt），可以獲得其他時間信號。信號還通過快速傅立葉變換（FFT）在頻域中映射，針對兩個向量的冪進行了優化（請參見 *Table 2.*）。

從第*II B* 節中描述的每個採樣窗口中，獲得特徵向量。特徵映射採用了以前在HAR文獻[2]中使用的標準度量，例如平均值，相關性，信號幅度區域（SMA）和自回歸係數[3]。為了提高學習性能，還採用了一組新的功能，包括不同頻帶的能量，頻率偏度和矢量之間的角度（例如，平均身體加速度和y矢量）[1]。表3列出了適用於時域和頻域信號的所有度量的列表。

最後，數據集的特徵工程部分由10299條具有561個特徵的記錄組成。然後將數據集進一步以70％到30％的比例隨機分為訓練和測試集。

| ***Table 2. List of Raw Signals in Time and Frequency Domains*** | | | |
| --- | --- | --- | --- |
| Signal-Dimension | Domain | Signal-Dimension | Domain |
| tBodyAcc-XYZ | Time | fBodyAcc-XYZ | Fequency |
| tGravityAcc-XYZ | Time | fBodyAccJerk-XYZ | Fequency |
| tBodyAccJerk-XYZ | Time | fBodyGyro-XYZ | Fequency |
| tBodyGyro-XYZ | Time | fBodyAccMag | Fequency |
| tBodyGyroJerk-XYZ | Time | fBodyAccJerkMag | Fequency |
| tBodyAccMag | Time | fBodyGyroMag | Fequency |
| tGravityAccMag | Time | fBodyGyroJerkMag | Fequency |
| tBodyAccJerkMag | Time |  |  |
| tBodyGyroMag | Time |  |  |
| tBodyGyroJerkMag | Time |  |  |

# III. 數據視覺圖

1. *Class and subject balance. Figure 1.* 我們加載每個參與者和活動類別的標籤，併計算每個參與者和活動類別之間的記錄數。每個科目和每個活動類別之間的平衡是均勻分佈的。根據此視覺圖本數據集並沒有類別不平衡的問題。

2) *Raw accelerometer signal in one time window*. *Figure 2.* 一個時間窗口中的加速度計信號如圖2所示。對於“ 躺*”* 活動，由於其幅度較小，因此可以自行分離。而其他活動類型將在頻域 (Fequency Domain) 中得到更好的分析。

3) *PCA + t-SNE. Figure 3.* 為了可視化特徵工程數據集，使用了PCA 和t-Distributed Stochastic Neighbor Embedding (t-SNE) [4] 降維技術。應用T-SNE技術使可視化高維數據集的距離變得更加容易。 根據此視覺圖我們大約可以判斷一般的學習模型對於較於靜態的動作如 ”站“ 和 ”坐“ 這兩個動作有可能會有辨認上個困難。而對於動態的動作如 ”走“ ，”上階梯“ 和 ”下階梯“ 會可以做出很好的區分。

|  |
| --- |
|  |
| ***Fig 1.*  Activities *s*ample counts for each subject** |

|  |  |
| --- | --- |
|  |  |
| ***Fig. 3 Sample counts for each activity class*** | |
|  | |
| ***Fig 5. t-SNE visualization of data for different activity class*** | |

# IV. 方法

1. 機器學習模型

對於特徵工程數據集上的分類任務，選擇了8個模型。這些模型是SVM [12], Extra Tree, Gradient Boost Machine, Random Forest [13], K-Nearest Neighbors, Bagging Ensemble, Decision Tree, 及 Naive Bayes.

1. 深度學習模型

在這份報告中，筆者選擇兩種基於卷積神經網絡（CNN）的基礎架構，RESNET 和FCN，作為深度學習模型來處理原始信號數據集。儘管[5] [7]進行的研究不是專門針對人類活動識別而設計的，但該模型是針對多元時間序列分類設計的，也已可以適用於處理ＨAR問題。此外，這些模型已經在各種不同的時間序列分類數據集中進行了測試和比較。在他們的研究中，Resnet和FCN 當前是TSC任務的表現最佳的模型。

|  |
| --- |
|  |
| ***Fig 2. Raw accelerator signal in X for one window in time domain*** |

由[6]調查顯示，已經有越來越多的深度學習模型被使用在HAR問題之上。對同一數據集的一項研究 [10] 也利用了另一種形式的CNN模型取得了與本研究相似的精準度。

在基於CNN的深度模型可以用於HAR任務之前，需要考慮以下幾個要求：

1）*Input adaptation*：與通常為3D（H xW x C）的圖像不同，大多數HAR任務由產生時間多維1D讀數（T x C）的傳感器訊號組成。因此，原來形式的CNN無法可以馬上適用於處理HAR問題。某種形式的輸入的轉換是必要的。如[6]所示，這種轉換主要有兩種類型：數據驅動和模型驅動。對於數據驅動的方法，每個維度都被視為一個渠道。使用模型驅動的方法，可以將輸入調整大小以變成虛擬2D圖像。對於Resnet和FCN模型，我們都使用數據驅動的方法。

*2*）*Pooling*：對於CNN，pooling 層通常用於避免 overfitting 及減少用於訓練的計算資源[8]。但是，由於與基於圖像的數據集相比，輸入數據集非常小，因此本實驗沒有使用 pooling。

3）W*eight sharing*：由於卷積層中的重量共享，CNN通常比DNN快。

最後，從淺度機器學習中選擇三個表現最好的模型來處理原生訊號。選擇它們是為了比較淺層模型和深層模型之間的性能。

1. *Resnet*

Residual Network (Resnet). 是在每個 residual 塊中通過快捷連接而連接的非常深層的神經網絡的架構。這些連接使梯度流可以直接流經底層。這樣做是為了減少非常深的網絡的“ 消失梯度” 問題。它是在 object detection 和其他與視覺相關的任務中實現 state of art 成果的架構之一[8]。

B. FCN

全卷積網絡（FCN）由於其效率和質量而顯示出對 semtenic segmentation 應用上使分強大 [11]。在HAR設置中，它主要用作特徵提取器，最終輸出仍然是 softmax 層。FCN通常被構造為堆棧數個卷積塊。而每個卷積塊由一個卷積層，batch normalization 層以及最後的RELU激活層組成。

| ***Fig 4. Architecture of Deep Learning models used in TSC and HAR*** |
| --- |
|  |
|  |

# IV. 實驗結果

1. 基於特徵工程數據的機器學習模型

在這些模型當中, 以 Linear SVM 表現在佳, Ensemble 其次。Naive Bayes 由於模型太過簡單而無法正確的區別三種較為靜態的活動 （坐, 躺, 站）。

| **Fig. 5. Confusion Matrix of Shallow Models On Feature Engineered Dataset** | | |
| --- | --- | --- |
|  | **Linear SVM** | **Extra Tree** |
|  |  |  |
|  | **GBM** | **Random Forest** |
|  |  |  |

|  |  |  |
| --- | --- | --- |
|  | **KNN** | **Bagging** |
|  |  |  |
|  | **Decision Tree** | **Naive Bayes** |
|  |  |  |

B. 基於原始信號數據的深度學習和機器學習

原始數據集還包括一組經過簡單前置處理的原始信號數據。原始信號數據集也是使用70:30的比例進行分割為 training 和 test。*Section A.* 中的三個最佳性能模型。用於對這些原始信號數據進行實驗。對於深度學習模型，使用 Resnet 和 FCN 模型。

1) *Ensemble Classifiers.* 與原始信號數據集上的其他淺層模型相比，Ensemble 分類器通常獲得更好的準確性得分。estimator 的數量也從原來的500增加到1000，以試圖更好地從原始信號中捕獲特徵。在這兩個淺層模型中，GBM的性能優於Extra Tree。

2) *SVM*. 線性核 (“linear” kernel) 在原始信號數據集上表現較差，這可能是因為線性核的模型容量不足以捕獲原生序號中的特徵。因此，kernel 被更改為“ rbf ” 。

| **Fig. 5. Confusion Matrix of Shallow Models On Raw Signal Dataset** | | |
| --- | --- | --- |
|  | **GBM** | **Extra Tree** |
| Ensembles |  |  |
|  | **SVM (rbf)** | **Linear SVM** |
| SVM |  |  |

3) 深度學習模型。由於數據集的數量以深度學習的標準來說並不大，所以很快就會 overfit。一定程度的 regularization 的處理會幫助其 validation accuracy

| **Fig 6. Confusion Matrix for Deep Learning Models On Raw Signal Dataset** | | |
| --- | --- | --- |
|  | Resnet | FCN |
|  |  |  |

# V. 結論

*Table 3.* 列出了在特徵工程數據集上執行的各種機器學習模型的準確度。*Table 4.* 列出了淺度和深度的機器學習模型在處理原生訊號的準確性。

| *Table 3. Accuracy on Featured Engineered Dataset* | |  | *Table 4. Accuracy on Raw Signal Dataset* | |
| --- | --- | --- | --- | --- |
| **Models** | **Accuracy** |  | **Models** | **Accuracy** |
| Linear SVM | 96.233 |  | Resnet | 95.215 |
| SVM (rbf) | 95.656 |  | FCN | 93.824 |
| Extra Tree | 94.706 |  | GBM | 90.77 |
| GBM | 94.265 |  | Extra Tree | 87.037 |
| Random Forest | 92.331 |  | SVM (rbf) | 76.956 |
| KNN | 89.583 |  | Linear SVM | 60.77 |
| Bagging | 84.323 |  |  |  |
| Decision Tree | 80.8282 |  |  |  |
| Naive Bayes | 57.142 |  |  |  |

SVM使用“ 線性” 內核的來處理已經經過特徵工程處理的的數據集時取得了最佳的精度，但對於處理原始信號數據集時其性能便大大降低。如果將內核切換為“ rbf ” ，性能將恢復約17％。但是，它的結果仍然遠非理想值。在淺層模型中，GBM在原始信號數據集方面表現最好。然而，在對於原生訊號的處理上所有的淺學習模型的表現仍比不上兩個深度學習模型。

深度學習模型的另一個優點是其準確性將隨訓練數據的數量增多而上升。在可見的未來隨著智能手機和智能手錶與運動手環的普及，用於活動識別任務的數據可用性將變得越來越高。深度學習的另一個特點是轉移學習 [14]。轉移學習可以利用大數據訓練一個基礎 backbone 模型。在使用者端可以利用該 backbone 所取得的特徵擷取能力訓練一個個人化的模型來適應每個使用者的獨特性。

但是，深度模型也不是完全沒有缺點。與淺層學習模型不同，Resnet等深層模型的訓練時間要多得多。在具有Nvidia 1080 GPU的工作站上，100個世代的訓練時間約為10分鐘。另一方面，在6核Intel Core i7 3770K工作站上使用淺層模型進行的平均培訓時間不到2分鐘。

由於傳統的淺層學習方式在處理已經普及化的動作感應器及個人化的需求上會有越來越多的限制，對於可以處理大量數據及讓使用者自行訓練而不用依靠專家知識的演算法的需求也會越來越高。甚至，如果想要完成環境感知運算的目標，淺層學習方式仍遠遠不夠。深度學習在此應用上正符合未來的趨勢。

# 參考文獻

[1] Davide Anguita, et al. “A Public Domain Dataset for Human Activity Recognition Using Smartphones.” 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.

[2] Yang, Jhun-Ying, et al. “Using Acceleration Measurements for Activity Recognition: An Effective Learning Algorithm for Constructing Neural Classifiers.” Pattern Recognition Letters, vol. 29, no. 16, Dec. 2008, pp. 2213–20. DOI.org (Crossref), doi:10.1016/j.patrec.2008.08.002.

[3] A.M. et al. Human activity recognition via an accelerometer- enabled-smartphone using kernel discriminant analysis. In *Proceedings of the 5th International Conference on Future Information Technology*, pages 1–6, 2010.

[4] L.J.P. van der Maaten and G.E. Hinton. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research* 9(Nov):2579-2605, 2008.

[5] Ismail Fawaz, Hassan, et al. “Deep Learning for Time Series Classification: A Review.” Data Mining and Knowledge Discovery, vol. 33, no. 4, July 2019, pp. 917–63. DOI.org (Crossref), doi:10.1007/s10618-019-00619-1.

[6] Wang, Jindong, et al. “Deep Learning for Sensor-Based Activity Recognition: A Survey.” Pattern Recognition Letters, vol. 119, Mar. 2019, pp. 3–11. arXiv.org, doi:10.1016/j.patrec.2018.02.010.

[7] Wang, Zhiguang, et al. “Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline.” ArXiv:1611.06455 [Cs, Stat], Dec. 2016. arXiv.org, http://arxiv.org/abs/1611.06455.

[8] He, Kaiming, et al. “Deep Residual Learning for Image Recognition.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2016, pp. 770–78. DOI.org (Crossref), doi:10.1109/CVPR.2016.90.

[9] Sousa Lima, Wesllen, et al. “Human Activity Recognition Using Inertial Sensors in a Smartphone: An Overview.” Sensors, vol. 19, no. 14, July 2019, p. 3213. DOI.org (Crossref), doi:10.3390/s19143213.

[10] Jiang, Wenchao, and Zhaozheng Yin. “Human Activity Recognition Using Wearable Sensors by Deep Convolutional Neural Networks.” Proceedings of the 23rd ACM International Conference on Multimedia - MM ’15, ACM Press, 2015, pp. 1307–10. DOI.org (Crossref), doi:10.1145/2733373.2806333.

[11] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[12] Davide Anguita, et al., “Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine,” in Ambient Assisted Living and Home Care, 2012.

[13] Toma At’s Peterek, et al., “Comparison of classification algorithms for physical activity recognition,” in Innovations in Bio-inspired Computing and Applications, 2014.

[14] Pan, Sinno Jialin, and Qiang Yang. “A Survey on Transfer Learning.” IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, Oct. 2010, pp. 1345–59. DOI.org (Crossref), doi:10.1109/TKDE.2009.191.