Title
Children Activity Recognition: Challenges and Strategies

Authors
Hosseini, Anahita
Fazeli, Shayan
van Vliet, Eleanne
et al.

Publication Date
2018

Peer reviewed

Children Activity Recognition: Challenges and Strategies

Anahita Hosseini[1], Shayan Fazeli[1], Eleanne van Vliet[2], Lisa Valencia[2],
Rima Habre[2], Majid Sarrafzadeh[1], Alex Bui[1]

Abstract— **In this paper, we study the problem of children activity recognition using smartwatch devices. We introduce the need for a robust children activity model and challenges involved. To address the problem, we employ two deep neural network models, specifically, Bi-Directional LSTM model and a fully connected deep network and compare the results to commonly used models in the area. We demonstrate that our proposed deep models can significantly improve results compared to baseline models. We further show benefits of activity intensity level detection in health monitoring and verify high performance of our proposed models in this task.**

1

I. INTRODUCTION

With the recent prevalence of smartwatches and smartphone technologies, wireless health systems and mobile health (mHealth) applications are increasingly starting to adopt these technologies for healthcare applications. As the popularity of smartwatches for health monitoring grows, so do the challenges that come with finding meaning in the newly available physiological data. One goal of many existing studies is to use data such as accelerometer and gyroscope readings for activity recognition. While great advances have been made in this area [1], [2], it remains an active area of research, partly because of the complexity and diversity of human movement based on age, health condition, and behavioral patterns.

Following our study aimed at prediction and prevention of asthma attacks in children [3] and clinical studies showing the impact of activity level on asthma exacerbation [4], an activity recognition model for children seemed necessary. However, the major focus of most activity recognition studies has been on adults [2], [5], [6], and the classic machine learning models used for adult activity recognition often do not translate well to children [7]. This is partly because of significant differences in the way children and adults perform basic activities such as running or climbing stairs.

In general, activity recognition imposes two main challenges specific to children:

1) Collecting large labeled datasets for children is difficult. Partly due to the fact that children tend to change activities more frequently and listen less well to the experimenter's commands. This is while in most activity recognition studies, class labels are obtained in constrained laboratory settings. [8].
2) There tend to be large variations between children when performing the same activities, which translates to more variation in the signal data being processed for activity recognition.

When studying health monitoring solutions for children, smartwatches show two main advantages over smartphones. Firstly, unlike smartphones that are bulky to handle during different activities, they allow for continuous data monitoring throughout the day and can easily be worn even in high levels of activity. This is especially important for applications such as asthma management. Secondly, they can collect additional data such as heart rate and externally connected sensor data for health monitoring purposes which are particularly promising in applications that require multiple sensor data for remote health monitoring and management.

In this study, we aim to tackle the smartwatch based children activity recognition problem. For this aim, we collected labeled activity data from 25 children aged 8-14 and compared the performance of state of the art deep neural network models to the widely used classic activity recognition models.

We show that mentioned challenges and limitations of smartwatches introduce major error to basic models and demonstrate that using a bidirectional recurrent neural network (BRNN)[9] can improve the results compared to other baseline models while not adding too much complexity as fully connected deep neural network models. We then show the results of our proposed models when used to capture the intensity of activity level with high accuracy, which is essential for remote monitoring of children and our efforts in asthma attack prevention.

To the best of our knowledge, this is the first study with a systematic focus on smartwatch-based children activity recognition. Results of this study can pave the way for other children health monitoring platforms that are dependent on children activity recognition.

II. Related work

A. Activity Recognition

A wealth of research has been dedicated to using accelerometer and gyroscope data for activity recognition [1][2][5]. A large amount of this work has gone into finding the optimal classifier for activity detection. Classic models, such as random forest, have shown some success, as shown by Casale et al.[1], who achieved between 90% and 94% accuracy in identifying activities such as walking, sitting,

[1]This research was supported by National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award number U54 EB022002 and conducted with IRB approval with code 15-001402.
1. Author is with University of California Los Angeles
2. Author is with University of Southern California - Los Angeles

and taking the stairs. Similarly, Qian et al.[2] used a simple SVM model on activity recognition data and achieved 89%. Recently, deep learning models have been applied to activity recognition. Yang et al.[6] showed that a convolutional neural network (CNN) can achieve up to 5% higher accuracy than SVM, which was used as the baseline. Ordez et al. [5] then compared recurrent neural networks (RNN) with other deep learning models and found that RNNs outperformed non-recurrent neural networks by an average of 4%. However, both of these studies have used multiple inertial sensors that are placed in different parts of the body which is not feasible on a daily basis. especially for children. Moreover, none of the above-mentioned studies have focused on children activity recognition.

B. RNN

Recurrent neural networks are a class of deep neural network models that employ memory concept or cyclical design and have shown significant performance improvements in the classification of sequential data, such as natural language processing and speech recognition[10], [11]. In past years different designs of memory cells and architectures have been proposed for RNNs among which Long short-term memory cells (LSTMs)[12] and bidirectional architecture for RNN (BRNNs)[9] has proven great performance in diverse domains. The LSTM memory cell, allows the model to better learn long-term dependencies. Graves and Schmidhuber[13] showed capabilities of LSTM as a powerful tool for recognizing unconstrained handwriting, such as cursive or Arabic words. Moreover, BRNN passes activation both backward and forwards in time, allowing the model to use inputs from the past and future to classify the current time frame. Graves, Mohamed, and Hinton [14] showed that a BRNN could be combined with the LSTM model to be applied to speech recognition and found that it produced the smallest known error in phoneme recognition. The design was further employed in many speech recognition studies reviewed in [15].

III. Data Collection

To collect children activity data, an Android smartwatch app was designed to record accelerometer and gyroscope signals in real-time and transfer the anonymized data to a web server for activity prediction. Figure 1 illustrates the overview of our activity recognition system. In this study, 25 children (10 girls and 15 boys) aged 8 to 14 years old were recruited to participate. Each child was asked to wear the smartwatch and perform six different activities as instructed. Activities included running, walking, standing, sitting, lying down, and stair climbing and each was recorded for a duration of 10 minutes and sensor data was collected with the frequency of 10 Hz. After instructing the children before each 10-minute time span, they were left free to perform activities to obtain a real-world data. Data collection was stopped during time spans in which children stopped doing the required activity.

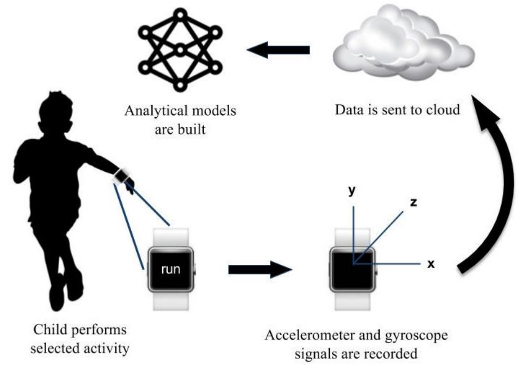

Fig. 1. The overall system architecture for a wearable activity monitoring device.

IV. Methodology

In this section, we review the data processing and machine learning strategies used in this study.

A. Data Representation

Representing raw signals received from sensors in an informative way is a necessity for achieving better interpretations of data. Although deep machine learning models allow for automated representation learning [16], [17], they ask for a tremendously large amount of labeled data, which is not available for children activities.

Time-series signals can be described nicely by their time-domain and frequency domain features. This enables us to efficiently work with small datasets for training high-performance inference models, without the need for heavy automatic representation learning.

Therefore, we first employed time-based windowing technique as it has shown superior performance compared to other methods [18] to segment the signal. Next, informative statistical features were extracted for each time-window to form its representation. We extracted widely used features for time-series analysis, studied in the literature [19].Table I lists the features extracted in this study from each or couple of axes of accelerometer and gyroscope signals.

TABLE I

Features extracted for each window in pre-processing stage

Feature Type	Axis
Every Single Axis	Mean,Median, range, min, max, std,25 and 75 percental RMS,zero crossing, fft-entropy Range, Integration
Every Two Axes	Correlation, delta
All Axes	Signal Vector Magnitude

B. Deep Models

Deep neural network models have shown great performance in information discovery and have outperformed classic and shallow models in learning hidden relations in different areas [20]. We study and compare two deep models one with sequential processing design(RNN) and one with a deep fully connected design.

1) Fully Connected Design: Once our signal is segmented into feature representations (windows), each window coupled with a label can be viewed as a data sample to be fed into the prediction models. To study the performance of deeper models over shallow ones, we designed a multi-layer neural network model in which layers were decoupled from each other by RELU nonlinear activation functions. As mentioned earlier, activity recognition datasets for children are usually small and more complex deep models cannot be trained on them. The best model found was a three-layer neural network with 100 neurons in each hidden layer. The model was trained using the Adam optimization scheme to help with the convergence.

2) Sequential Design: In another approach, instead of working with one single window at a time, we can view extracted representations as a consecutive window sequence and learn from activities happening close to a time-window. For this aim, a Bi-Directional Recurrent Neural Network, specifically a Bi-directional LSTM network, was employed. Such model can receive a sequence of input activity windows of any length and in each stage make the prediction based on the input in that stage and the memory state that is formed by processing the previous windows.
LSTM [12] configuration was employed to tackle the problem of vanishing gradient and enable the network to better recognize what information is worth remembering, and what is not. Moreover, Bi-directional architecture enabled the model to process the sequence of inputs in both forward and backward directions to better capture time dependencies for each activity.
The structure of the designed BiLSTM network is shown in Figure 2. In our design, an LSTM layer in which each unit has 32 neurons is followed by two layers of the fully connected network with 50 and 25 neurons and prediction is done through a final softmax layer.

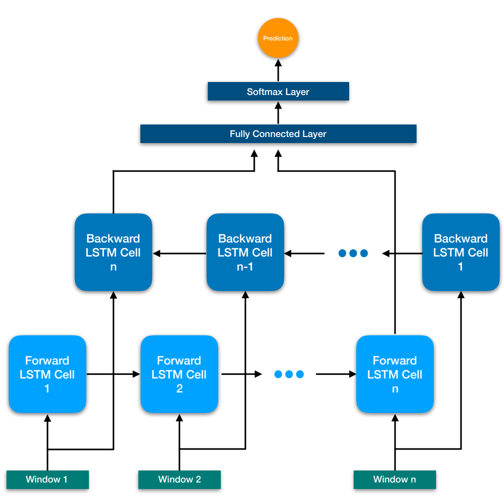

Fig. 2. Bi-Directional LSTM, along with fully connected and softmax layer to predict the activity based on the sequence of time-window representations

V. Results

To validate the performance of our proposed deep models, we chose two widely used models in the area, random forest(RF) and a shallow one layer neural network(FF1) as our baselines. Both models were tested over validation sets to find the best configuration which was training random forest with 50 decision trees and neural network with 100 RELU nodes. Figure 3 shows the F1-score of all models for prediction of each activity. It can be easily inferred that RNN model shows higher performance than the other deep model (FF3) and baseline models. FF1 although shows strong results on detection of walk activity, it cannot almost detect sitting or standing activities. Random Forest also shows competitive results with deep models in walking and running tasks, however, performs poorly on other tasks compared to deeper models. We can also see that the deep network (FF3) achieves competitive results to RF, however, cannot beat the RNN model in most cases. One important observation is that the obtained results from children show generally lower accuracy than reported results for adults [2] due to high variance in their activity shape. To study the

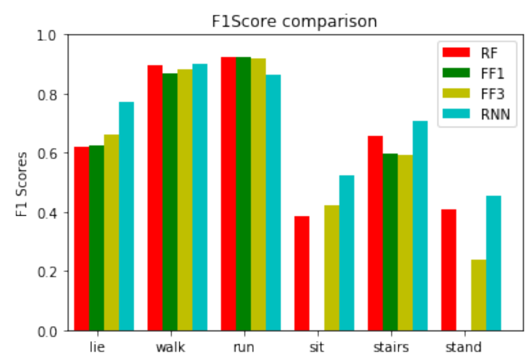

Fig. 3. F1-score results of activity prediction comparing deep models to baseline models

source of accuracy loss in previous results we analyze the confusion matrix of activities in Figure 4. We can infer that the model has difficulty in distinguishing sitting and standing. This confusion is due to the fact that movement of the wrist in these two activities can be very similar to each other especially when the hand is left free to the side. The same confusion rise for lying and standing activities. These challenges are inevitable when smartwatches are used as the activity recognition tracker. As discussed earlier, in many applications of activity recognition on children, such as asthma exacerbation prevention, the intensity level of activity is of greater importance than the activity itself. Figure 5 demonstrates the performance of our model when classifying activities into 3 levels of low, medium, and high intensity. RNN model allows us to detect intensity level of activities in children with a more than 80 percent of the average F1 score.

Overall, results of our experiments prove the superiority of recurrent neural network detection of children activity.

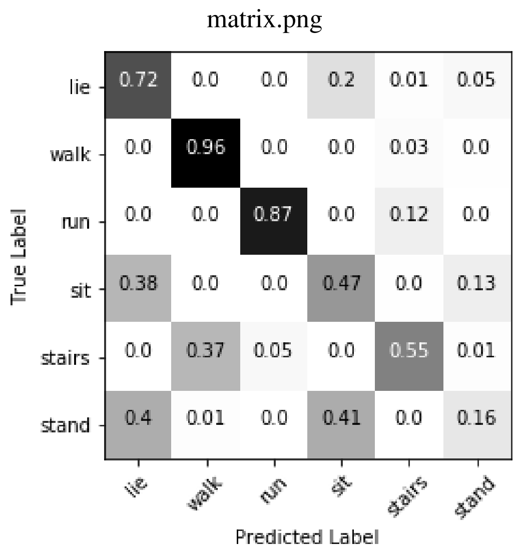

Fig. 4. Confusion matrix for six detected activities

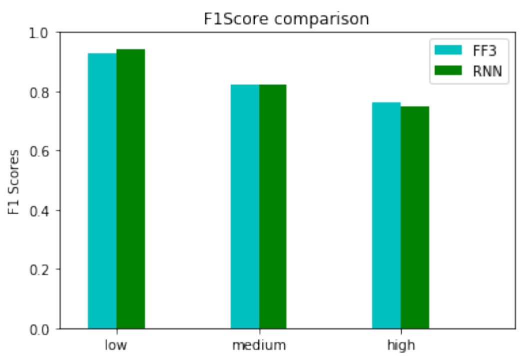

Fig. 5. F1-score performance on activity intensity level prediction

VI. CONCLUSIONS

This study focused on the problem of children activity recognition, which is of great importance in the domain of remote health monitoring for children. We showed that variance in activity and limitations of smartwatches are major challenges for this task. We also demonstrated that RNN based models can be a good choice for children activity recognition because of their ability to capture more information while being simple enough to be trained on small datasets. Future work in this area should focus on personalized learning of activity for children. Learning a model for a group of children cannot translate well into each individual. However, if models adapt to the activities of each child, results can be improved.

REFERENCES

[1] P. Casale, O. Pujol, and P. Radeva, "Human activity recognition from accelerometer data using a wearable device," *Pattern Recognition and Image Analysis*, pp. 289–296, 2011.

[2] H. Qian, Y. Mao, W. Xiang, and Z. Wang, "Recognition of human activities using svm multi-class classifier," *Pattern Recognition Letters*, vol. 31, no. 2, pp. 100 – 111, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865509002530

[3] A. Hosseini, C. M. Buonocore, S. Hashemzadeh, H. Hojaiji, H. Kalantarian, C. Sideris, A. A. Bui, C. E. King, and M. Sarrafzadeh, "Feasibility of a secure wireless sensing smartwatch application for the self-management of pediatric asthma," *Sensors*, vol. 17, no. 8, p. 1780, 2017.

[4] D. Koinis Mitchell, S. J. Kopel, C. A. Esteban, R. Seifer, N. W. Vehse, S. Chau, and E. Jelalian, "Asthma status and physical activity in urban children," in *B23. NOVEL EPIDEMIOLOGY, MANAGEMENT, AND OUTCOMES IN ASTHMA*. Am Thoracic Soc, 2017, pp. A2993–A2993.

[5] F. J. Ordez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, 2016.

[6] J.-Y. Yang, J.-S. Wang, and Y.-P. Chen, "Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers," *Pattern Recognition Letters*, vol. 29, no. 16, pp. 2213 – 2220, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865508002560

[7] S. Suzuki, Y. Mitsukura, H. Igarashi, H. Kobayashi, and F. Harashima, "Activity recognition for children using self-organizing map," in *RO-MAN, 2012 IEEE*. IEEE, 2012, pp. 653–658.

[8] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," *Pervasive computing*, pp. 1–17, 2004.

[9] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional lstm and other neural network architectures," *Neural Networks*, vol. 18, no. 5, pp. 602–610, 2005.

[10] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[11] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*. IEEE, 2013, pp. 6645–6649.

[12] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[13] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," in *Advances in Neural Information Processing Systems 21*, D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, Eds. Curran Associates, Inc., 2009, pp. 545–552. [Online]. Available: http://papers.nips.cc/paper/3449-offline-handwriting-recognition-with-multidimensional-recurrent\-neural-networks.pdf

[14] A. Graves, A. r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 6645–6649.

[15] A. Zeyer, P. Doetsch, P. Voigtlaender, R. Schlüter, and H. Ney, "A comprehensive study of deep bidirectional lstm rnns for acoustic modeling in speech recognition," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2462–2466.

[16] M. A. Alsheikh, A. Selim, D. Niyato, L. Doyle, S. Lin, and H.-P. Tan, "Deep Activity Recognition Models with Triaxial Accelerometers." in *AAAI Workshop: Artificial Intelligence Applied to Assistive Technologies and Smart Environments*, 2016.

[17] S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. Abdelzaher, "Deepsense: A unified deep learning framework for time-series mobile sensing data processing," in *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2017, pp. 351–360.

[18] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas, "Window size impact in human activity recognition," *Sensors*, vol. 14, no. 4, pp. 6474–6499, 2014.

[19] D. Figo, P. C. Diniz, D. R. Ferreira, and J. M. Cardoso, "Preprocessing techniques for context recognition from accelerometer data," *Personal and Ubiquitous Computing*, vol. 14, no. 7, pp. 645–662, 2010.

[20] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.