Journal of Science and Medicine in Sport 18 (2015) 426–431

Contents lists available at ScienceDirect

Journal of Science and Medicine in Sport

journal homepage: www.elsevier.com/locate/jsams

Original research

Prediction of activity type in preschool children using machine learning techniques

Markus Hagenbuchner[a], Dylan P. Cliff[b,*], Stewart G. Trost[c], Nguyen Van Tuc[a], Gregory E. Peoples[d]

[a] Faculty of Engineering and Information Science, University of Wollongong, Australia
[b] Faculty of Social Sciences, Early Start Research Institute, University of Wollongong, Australia
[c] Institute of Health and Biomedical Innovation, Queensland University of Technology, Australia
[d] School of Medicine, University of Wollongong, Australia

ARTICLE INFO

Article history:
Received 27 January 2014
Received in revised form 2 June 2014
Accepted 7 June 2014
Available online 24 June 2014

Keywords:
Physical activity
Pattern recognition
Accelerometry
Neural networks
Exercise
Validity

ABSTRACT

Objectives: Recent research has shown that machine learning techniques can accurately predict activity classes from accelerometer data in adolescents and adults. The purpose of this study is to develop and test machine learning models for predicting activity type in preschool-aged children.
Design: Participants completed 12 standardised activity trials (TV, reading, tablet game, quiet play, art, treasure hunt, cleaning up, active game, obstacle course, bicycle riding) over two laboratory visits.
Methods: Eleven children aged 3–6 years (mean age = 4.8 ± 0.87; 55% girls) completed the activity trials while wearing an ActiGraph GT3X+ accelerometer on the right hip. Activities were categorised into five activity classes: sedentary activities, light activities, moderate to vigorous activities, walking, and running. A standard feed-forward Artificial Neural Network and a Deep Learning Ensemble Network were trained on features in the accelerometer data used in previous investigations (10th, 25th, 50th, 75th and 90th percentiles and the lag-one autocorrelation).
Results: Overall recognition accuracy for the standard feed forward Artificial Neural Network was 69.7%. Recognition accuracy for sedentary activities, light activities and games, moderate-to-vigorous activities, walking, and running was 82%, 79%, 64%, 36% and 46%, respectively. In comparison, overall recognition accuracy for the Deep Learning Ensemble Network was 82.6%. For sedentary activities, light activities and games, moderate-to-vigorous activities, walking, and running recognition accuracy was 84%, 91%, 79%, 73% and 73%, respectively.
Conclusions: Ensemble machine learning approaches such as Deep Learning Ensemble Network can accurately predict activity type from accelerometer data in preschool children.

1. Introduction

Due to the limitations of self-reports and pedometers, as well as the intermittent activity patterns of children, accelerometry has become the 'best-practice methodology' for assessing physical activity (PA) and sedentary behaviour in pre-schoolers, school-aged children and adolescents.[1,2] To interpret accelerometry count data, researchers have typically used cut-points developed from regression or receiver operating characteristic curve analyses to estimate time spent in sedentary behaviour, and light, moderate and vigorous intensity PA. However, conventional regression-based approaches are limited in their ability to accurately predict energy expenditure across a wide range of activities,[3–5] because the relationship between accelerometer counts and energy expenditure (EE) differs according to the type of activity performed. Not surprisingly, cut-point methods Exhibit 28–45% misclassification of PA intensity in children and adolescents.[3,5,6] As accelerometry use is widespread, this level of misclassification has significant implications for understanding and promoting PA among children and adolescents internationally.

Innovative data processing methodologies such as those utilising machine learning approaches, provide PA researchers with the potential to substantially improve the accuracy of PA measurement. Machine learning is an area of research concerned with the design and development of algorithms that allow computers

* Corresponding author at: Faculty of Social Sciences, University of Wollongong, Northfields Avenue, Wollongong 2522, Australia.
E-mail addresses: markus@uow.edu.au (M. Hagenbuchner), dylanc@uow.edu.au (D.P. Cliff), s.trost@qut.edu.au (S.G. Trost), vtn966@uow.edu.au (N. Van Tuc), greg_peoples@uow.edu.au (G.E. Peoples).

http://dx.doi.org/10.1016/j.jsams.2014.06.003
1440-2440/

to "learn" from data. The ability to recognise complex patterns and make intelligent decisions based on data is the main focus of machine learning research. An important class of machine learning algorithms is Artificial Neural Networks (ANN). ANNs are typically applied to applications where the complexity of the data or the task makes the design of alternative approaches impractical.

To date, just two studies have employed ANNs to predict activity type in children and adolescents. Trost and colleagues[6] developed and tested an ANN to classify PA type from second-by-second hip-worn ActiGraph data in 5–15 year-olds. Participants completed 12 activity trials that were categorised into 5 activity types: sedentary, walking, running, light intensity house-hold activities or games, and moderate-to-vigorous games or sports. Mean accuracy for activity type ranged from 81.3% to 88.4%. De Vries et al. trained an ANN to predict 9–12 year old children's PA type from accelerometers worn on the hip and ankle.[7] The overall classification accuracy across the seven activity types evaluated ranged from 57.2% (GT1M/ankle placement) to 76.8% (GT3X/hip placement).

Although the aforementioned studies indicate that machine learning approaches are feasible and offer enhanced accuracy for accelerometry-based assessments of PA in school-aged children and adolescents, the validity of neural networks developed in preschool-aged children has not been investigated. Due to developmental, biomechanical, and behavioural factors, such as differences in motor proficiency,[8] and PA types and patterns,[1,9] models developed in older children might not be generalisable to young children. To our knowledge, machine learning based accelerometry data modelling approaches are yet to be evaluated in pre-school children. Furthermore, previous models developed in school-aged children and adolescents have been trained and tested using conventional feed-forward ANNs with a single hidden layer, also known as Multi-Layer Perceptron Networks (MLP). Therefore, this study aimed to examine and compare the accuracy of MLP as well as more advanced models, such as a deep-learning-inspired neural network, for predicting PA type in preschool children.

2. Methods

Eleven children aged 3–6 years (mean age = 4.8 ± 0.87; 55% girls; mean BMI = $15.9 \pm 1.0\,\text{kg/m}^2$, 9.1% overweight)[10] were recruited to participate in the study via University staff email lists and word-of-mouth. Parent consent was obtained prior to participation. The study was approved by the University of Wollongong Human Research Ethics Committee.

Participants completed 12 structured activity trials (see Supplementary Table for a description of each activity) over two laboratory visits scheduled within a 3-wk period. Participants undertook the following six trials at visit 1: watching TV (TV), sitting on floor being read to (reading), standing making a collage on a wall (art), walking (walking), playing an active game against an instructor (active game), and completing an obstacle course (obstacle course). The remaining six trials were completed at visit 2: sitting on a chair playing a computer tablet game (tablet), sitting on floor playing quietly with toys (quiet play), treasure hunt (treasure hunt), cleaning up toys (clean-up), bicycle riding (bicycle), and running (running). Each trial was completed for 4–5 min. These 12 activities were then grouped into five activity classes: sedentary activities (TV, reading, tablet, and quiet play), light activities and games (art, treasure hunt, and clean-up), moderate to vigorous activities (active game, obstacle course, and bicycle), walking, and running.

Supplementary Table related to this article can be found, in the online version, at doi:10.1016/j.jsams.2014.06.003.

Participants were fitted with an ActiGraph GT3X+ (ActiGraph, Pensacola, FL) on the mid-axillary line at the iliac crest. The GT3X+ records time varying accelerations ranging in magnitude from ± 6 g. The acceleration output is digitised by a 12-bit analogue-to-digital converter at a user-specified rate (30–100 Hz). A sampling frequency of 100 Hz was used in this study.

For each activity trial, 1 s count data between minutes 2 and 4 was used for analyses. Since each of the eleven participants performed 12 different activity trials, there were a total of 120 s *11 subjects *12 trials = 15,840 instances of data available for the experiments. The 120 s segment was divided into non-overlapping time windows. Window sizes of 10 s, 15 s, 20 s, 30 s, and **60 s** were evaluated (parameters in bold font indicate the optimal configuration). For each window, features were extracted from those data instances. For ease of comparisons we utilised the same features used by Trost and colleagues.[6] These included the 10th, 25th, 50th, 75th and 90th percentiles and the lag-one autocorrelation values.

Three different ANNs were evaluated in this study: the standard feed-forward Multi-Layer Perceptron Network (MLP), the Self-Organizing Map (SOM), and the Deep Learning Ensemble Network (DLEN). The MLP is a supervised learning model and commonly consists of three layers: input, hidden and output layers.[11] Neurons in those layers are fully connected by a set of adjustable parameters called "weights". These weights are updated by a learning function which requires an input (training) set consisting of numeric features and associated target values. Consequently, the number of neurons in the input and output layer must match the dimension of input samples and the dimension of class labels respectively. The dimension of the hidden layer can be adjusted freely. The schematic of the MLP is shown in Fig. 1(a).

The SOM is an unsupervised learning model that is popularly applied to tasks requiring dimension reduction or clustering.[12] The SOM is computationally very efficient which makes it particularly useful for data mining.[12] Fig. 1(b) depicts the schematic of the SOM. Both MLP and SOM take in inputs in the form of vectors. If those vectors are long in size, it refers to the high dimensional input/data space. The SOM can project its input vectors to a 2-dimensional grid referred to as the "activation map", such that each input vector is then represented by a 2-dimensional vector or low dimensional data.

Because the MLP tends to perform poorly when dealing with limited number of samples and high dimensional input space, it makes sense to combine the SOM with MLP since they have complementary properties. The SOM has advantages over the MLP in that the algorithm is trained unsupervised. The resulting model is much less sensitive to "noise" or variability in the data. The MLP on the other hand is trained supervised, and has good generalisation properties. Therefore, adopting concepts from Deep Learning,[13] we evaluated the performance of the ensemble model DLEN consisting of a SOM as a first layer, followed by an MLP as a second layer. Both layers were trained on the same set of data with the second layer receiving the output of the first layer as an additional input.

The MLP and SOM models were implemented in plain C programming language. The SOM's parameters including the learning rate was selected from 0.6, 0.8, **1.0**, 1.2 and the radius in 12, 15, 20, **25**. The SOM activation map sizes tried were 19×17, $\mathbf{20 \times 19}$, 23×20 and 25×22. A number of MLP configurations were decided by assigning the size of the hidden layer to 3, 8, 13, 17 or **25** and the learning rate to 0.001, **0.01** or 0.5. For each validation round, the MLP and SOM were evaluated 10 times using different random initial conditions. The trained models providing the best performance on the training set was selected to produce the result for the test set. Both the MLP and SOM were trained for 10,000 iterations.

The leave-one-subject-out cross validation approach was used for model assessment. Thus, the model was trained on all input samples except for the data of one participant as the test set. After training, the model was then tested on the left-out data. The experiment was repeated until each participant was considered exactly once for testing. For comparison purposes the MLP results served