Activity Profiling and Phenotypes of Physical Activity and Sleep

Deepika Verma

Department of Computer Science
Norwegian University of Science and Technology, Trondheim, Norway
http://www.idi.ntnu.no

Abstract. Physical inactivity and obesity contribute significantly to several health problems and have been a topic of research amongst the public health research community. Objective measurements of physical behaviour using body-worn sensors have radically uplifted the method of collecting physical behaviour data from a large population and opened the possibility of analysing these objective recordings using machine learning methods to gain useful insights such as identifying different physical activities and the duration of time spent in each activity, among others. We aim to address the theme of physical activity from both the public health and computer science perspective by first determining the best strategy to increase the accuracy of human activity classification from body-worn sensor data collected in cohort studies using the same data collection protocol as HUNT4[1] and strategically using objectively measured physical activity data to identify clusters of different physical behaviour phenotypes such that intra-cluster similarity is maximized. Identifying physical behaviour phenotypes can allow us in the future to design tailored interventions for users to self-manage their daily physical activity routines. Additionally, we consider the compositional nature of the physical behaviour data we have and plan to employ suitable methodologies which allow us to model the compositional co-dependency between physical behaviours and sleep.

Keywords: Case-Based Reasoning, Machine Learning, Activity Phenotypes

1 Introduction

Maintaining an active lifestyle has been shown to be essential in leading a healthy life and has several positive effects such as improved mental health [4], sleep quality [6] and reduced risk of mortality [1]. Physical inactivity has been associated with several health problems in addition to negative impacts on mental health and mortality. Objective measurements of physical behaviour have opened new possibilities of research for public health and computer science researchers alike. Furthermore, it facilitates providing activity recommendations to the users of

[1] https://www.ntnu.no/hunt4

smart-activity trackers (such as FitBit) by keeping track of their total amount of physical activity. However, the activity recommendations remain nearly unchanged throughout the user base and may or may not be challenging enough for certain users since they are not tailored to suit each user's daily physical activity routine. To provide more tailored and personalized activity recommendations, accurate identification of different physical behaviours from body-worn sensors is a prerequisite.

In addition to regular physical activity, sleep duration and sleep quality are important but highly volatile factors for sound health and have a strong compositional co-dependency with daily physical behaviour. Although there is no doubt that physical behaviour and sleep have a profound impact on health, it has recently become evident that our current understanding of the health effect of these entities is flawed by serious methodological shortcomings. In particular, this pertains to the use of self-reported data, which is prone to bias and misclassification [7], and more importantly, the failure to recognise the compositional nature of various physical behaviours and sleep [2,3]. The latter refers to the fact that the duration of various physical behaviours and sleep are inherently co-dependent, that is, increasing the duration of one behaviour (for example, sitting) will necessarily reduce the duration of at least one other behaviour since the total time in any given day is fixed at 24 hours. Therefore, it is imperative to determine whether a change in one type of behaviour leading to compensatory shifts in other behaviours is beneficial or harmful for health.
This research builds on the requirement to generate classification models which are general enough to provide accurate classification of distinct physical behaviours from sensor data collected in cohort studies with the same data collection protocol as HUNT4 and furthermore, to effectively utilise the objective data to identify different physical behaviour phenotypes. HUNT4 is the fourth round of the HUNT[2] cohort study which collects large amount of both subjective as well as objective health data and objective physical activity data collected through body-worn accelerometers (which forms the target dataset in this research).

2 Research Aim

The overall goal is to achieve a high classification accuracy in classifying basic human activities from sensor data streams collected during HUNT4 and utilise this information to correctly estimate the time spent by an individual in each of the six different lower level physical activities: *lying, sitting, standing, walking, running, cycling*. Furthermore, these objective duration of physical activities of the participants are intended to be utilised to identify phenotypes of physical activities, which is important in order to design personalised and sustainable digital interventions for users for self-managing daily physical activity routines.

[2] `https://www.ntnu.no/hunt/`

2.1 Foreseen challenges

- How to improve performance of machine learning classifiers for identifying human behaviour accurately using sensor data collected during HUNT4?
- How to identify sleep duration from multiple sensor data streams using machine learning methods?
- How to analyse the quality of clusters (physical behaviour phenotypes) obtained using different clustering methods?
- How to design tailored interventions of daily physical activity routines for users using existing physical behaviour profiles?

To address the first challenge, we will use different training datasets, including one wherein the data has been collected in out-of-lab settings and evaluate their performance by comparing with other model(s) trained using dataset(s) collected in in-lab settings. To address the second challenge, we will implement methodologies to identify sleep patterns and duration from Polysomnographic (sleep) recordings. Finally we plan to implement a similarity-based method for clustering the physical activity profiles of the HUNT4 participants. To analyse the quality of clusters obtained, we plan to implement state-of-the-art clustering methods and compare the performance of the similarity-based method with them. Furthermore, we plan to utilise clustering evaluation measures for comparing the cluster quality of the implemented methods to generate a quality assessment. The resulting clusters are expected to localise our search for the physical behaviour phenotypes in our dataset.

2.2 Expected Outcomes

Overall, the main aim of this thesis is the exploration of machine learning methods for the HUNT4 physical behaviour raw data recordings. We expect to build a toolbox that allows:

- Analysis of the body-worn accelerometer data using machine learning methods.
- Analysis of the high-resolution physical behaviour data using CBR and Compositional data analysis.

3 Proposed Research Plan

3.1 Literature Review

A thorough review of the existing state-of-the-art and literature of human activity recognition systems forms the basis of this research. Identifying gaps in the existing work enables us to make improvements on previously untouched topics. Additionally, an overview of the literature and methodologies for Compositional Data Analysis is required for understanding the co-dependency between various physical behaviours and their influence on one another since the data we are using is compositional in nature.

3.2 Machine Learning Approach

Several state-of-the-art methods such as Random Forest, k-NN, SVM and Hidden Markov Models among others have been successfully implemented for human activity recognition systems and have been shown to be quite effective in identifying different physical behaviour. Neural Network implementations such as Recurrent Neural Networks [8], Long Short-Term Memory [5], Convolutional Neural Networks among others have also shown promising results with an added advantage of not requiring the feature engineering step before feeding the data into the algorithms. We plan to implement a combination of these machine learning methods using both in-lab and out-of-lab training datasets and generate a comparative assessment with respect to the data collection protocol as well as the subject-dependency of the methods in order to determine the method(s) best suited to the application.

3.3 Public Health Insights with CBR and Compositional Data analysis

For gaining insights from the public health dataset we have, we plan to utilise different methodologies including CBR and Compositional Data analysis. CBR has been used in our work for developing the similarity measures for assessing the similarity amongst and finding similar physical behaviour profiles of the participants of HUNT4. Furthermore, we will utilize the *knowledge intensive similarity* used in CBR to generate clusters of semantically similar behaviour profiles in order to localise our search for the physical behaviour phenotypes. Additionally, we will employ methodologies for compositional data analysis to estimate the effect of increase or decrease in sleep and/or sedentary time on the composition of daily activities since an increase in the duration of one behaviour, say sitting, will necessarily reduce duration of at least one other behaviour because the total time in any given day is fixed at 24 hours. Thus, whether a change in one type of behaviour is beneficial or harmful for health depends on the compensatory shifts in other behaviours. This data analysis methodology will be studied extensively along with various other related research articles to get a better understanding of it and how it can be applied to gain insights into the public health data and thereby utilize it to design tailored and sustainable digital interventions for the users to self-manage their daily physical activity routines .

4 Current Progress and Future Work

We have presented our methodology for developing the similarity measures in myCBR for assessing the similarity amongst the physical behaviour profiles at ICCBR 2018 [10] and NAIS 2019 [9]. Additionally, we have now implemented a *knowledge intensive similarity*-based method for clustering the physical behaviour profiles in order to localise the most similar profiles into a single cluster. We plan to perform bout analysis on fine-grained sequential physical activity data of the participants to generate their unique profiles and subsequently apply similarity-based clustering to obtain more refined clusters. We expect these clusters to give us a representation of the physical behaviour phenotypes in our dataset. On the human activity classification side, we plan to implement some state-of-the-art machine learning algorithms using both in-lab and out-of-lab datasets and generate a comparative assessment in order to determine the best suited method(s) for our dataset.

References

1. Arem, H., Moore, S.C., Patel, A., Hartge, P., Berrington de Gonzalez, A., Visvanathan, K., Campbell, P.T., Freedman, M., Weiderpass, E., Adami, H.O., Linet, M.S., Lee, I.M., Matthews, C.E.: Leisure Time Physical Activity and Mortality: A Detailed Pooled Analysis of the Dose-Response Relationship. JAMA Internal Medicine 175(6), 959–967 (06 2015)
2. Dumuid, D., Pedisic, Z., Stanford, T.E., Martin-Fernandez, J.A., Hron, K., Maher, C.A., Lewis, L.K., Olds, T.: The compositional isotemporal substitution model: A method for estimating changes in a health outcome for reallocation of time between sleep, physical activity and sedentary behaviour. Statistical Methods in Medical Research 28(3), 846–857 (2019)
3. Fairclough, S.J., Dumuid, D., Taylor, S., Curry, W., McGrane, B., Stratton, G., Maher, C., Olds, T.: Fitness, fatness and the reallocation of time between children's daily movement behaviours: an analysis of compositional data. International Journal of Behavioral Nutrition and Physical Activity 14(1) (may 2017)
4. Harris, M.A.: The relationship between physical inactivity and mental wellbeing: Findings from a gamification-based community-wide physical activity intervention. Health Psychology Open 5(1), 2055102917753853 (2018)
5. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation 9(8), 1735–1780 (1997)
6. Kredlow, M.A., Capozzoli, M.C., Hearon, B.A., Calkins, A.W., Otto, M.W.: The effects of physical activity on sleep: a meta-analytic review. Journal of behavioral medicine 38(3), 427–449 (2015)
7. Prince, S.A., Adamo, K.B., Hamel, M.E., Hardt, J., Gorber, S.C., Tremblay, M.: A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. International Journal of Behavioral Nutrition and Physical Activity 5(1), 56 (Nov 2008)
8. Twomey, N., Diethe, T., Fafoutis, X., Elsts, A., McConville, R., Flach, P., Craddock, I.: A comprehensive study of activity recognition using accelerometers. Informatics 5(2) (2018)
9. Verma, Deepika, Bach, Kerstin, Jarle Mork, P.: Similarity measure development for case-based reasoning- a data-driven approach. arXiv.org (May 2019), `https://arxiv.org/abs/1905.08581`
10. Verma, D., Bach, K., Mork, P.J.: Modelling similarity for comparing physical activity profiles - a data-driven approach. In: Cox, M.T., Funk, P., Begum, S. (eds.) Case-Based Reasoning Research and Development. pp. 415–430. Springer International Publishing, Cham (2018)