Sedentary workers recognition based on machine learning

Sotiris Alexiou†
Department of Electrical and Computer Engineering
University of Patras
Patras, Greece
salexiou@ceid.upatras.gr

Nikos Fazakis†
Department of Electrical and Computer Engineering
University of Patras
Patras, Greece
fazakis@ece.upatras.gr

Otilia Kocsis
Department of Electrical and Computer Engineering
University of Patras
Patras, Greece
okocsis@bok.gr

Nikos Fakotakis
Department of Electrical and Computer Engineering
University of Patras
Patras, Greece
fakotaki@upatras.gr

Konstantinos Moustakas
Department of Electrical and Computer Engineering
University of Patras
Patras, Greece
moustakas@ece.upatras.gr

ABSTRACT

Sedentary Behavior (SB) is one of the most frequent human behaviors and is associated with a multitude of extreme chronic lifestyle diseases and premature death. Office workers in particular are at an increased risk due to their extensive amounts of occupational sedentary behavior. There are currently several large data sets available which can be used to create long-term health risk prediction models. However, in many cases, the relation between physical activity and SB in work environment, is missing. The main objective of our study was to develop a method for the automatic classification of sedentary and non-sedentary workers in English Longitudinal Study of Ageing (ELSA) database.

CCS CONCEPTS

• **Computing methodologies → Artificial intelligence.**

KEYWORDS

Sedentary behavior, Old workers, Classification, Machine learning

ACM Reference format:

Sotiris Alexiou, Nikos Fazakis, Otilia Kocsis, Nikos Fakotakis and Konstantinos Moustakas. 2020. Sedentary workers recognition based on machine learning. In *The 13th PErvasive Technologies Related to Assistive Environments Conference (PETRA '20), June 30–July 3, 2020, Corfu, Greece. ACM, New York, NY, USA, 2 pages.* https://doi.org/10.1145/3389189.3397654

†Both authors contributed equally to this work.

PETRA '20, June 30-July 3, 2020, Corfu, Greece

ACM ISBN 978-1-4503-7773-7/20/06…$15.00
https://doi.org/10.1145/3389189.3397654

1 INTRODUCTION

Sedentary behavior is described as any waking behavior that is characterized by low body energy consumption while in a sitting or reclining position. Longer durations of sedentary activity have been reported to contribute to increased obesity, type 2 diabetes and reduced bone density, cardiovascular diseases and upper body pain. According to recent studies, the amount of time spent on sitting is a highly important determinant of health and wellbeing [9].

Many large data sets are currently available, including ELSA [1] repository, which can be used to establish long-term health risk prediction models for such human behaviours. However, in many cases the information related to the physical activity level performed in work environment is missing. The SmartWork [6] project builds a Worker-Centric, AI System for work ability sustainability for older workers. Moreover, most of the SmartWork services, target to model and support self-management of chronic health conditions by older office workers, who usually have a sedentary lifestyle and poor physical fitness. For the initialization of virtual user models and work functions in the SmartWork project, group profiles of users were built using ELSA database. The goal of our study is to develop a method for the automatic classification of sedentary and non-sedentary workers, based on other attributes (e.g. demographics) of the participants in ELSA.

2 BACKGROUND AND METHODOLOGY

The cross-sectional association between sedentary lifestyle and older people was examined by multiple demographic factors (age, sex, race/ethnicity, current job, education level, years of education, income), health factors (self-health report, frequency of moderate/vigorous/lite physical activity, blood pressure, heart diseases, lung diseases, diabetes, cholesterol , body mass index) [4]. A strong positive cross-sectional relationship between Body Mass Index (BMI) and a sedentary lifestyle has been consistently observed in numerous studies [7]. In addition, it is proved that physical activity (PA) is an influencing factor for healthy aging and lack of PA has been associated with chronic diseases caused from SB [2],[5]. At the same time, Parry and Straker [8], claim that a greater proportion of the workforce is now employed in low activity occupations such as office work and the contribution of the overall sedentary exposure is independently associated with poor health.

The ELSA collects data from people aged over 50 in waves, which are set 2 years appart. More than 18,000 people have taken part in the study since it started in 2002, with the same people re-interviewed every two years. The constructed training and testing dataset, which is a subset of ELSA, was created by studing a large number of attributes that are correlated to and potentially relevant for the identification of SB. However, ELSA has more than 3,800 attributes and 7 different waves that were considered in our study. In order to establish a unified set of attributes for each user, we derived harmonized variables for each attribute, by combining all 7 waves according to predefined attribute rules (such as numeric averages and categorical majority votings). Moreover, the class was created based on ELSA's occupation physical effort feature. The class distribution is balanced containig in total 8,991 observations. In order to reduce the dataset dimensionality and complexity, we chose only the features which are relevant to the specific class by using a feature selection method which ranks all attributes with respect to the class correlation. More specific, we utilized a feature selection method based on a variation of Random Forests [3]. According to this method, the attributes are ranked using the gini importance score of the model's trees, we decided to keep the first 23 features with the highest cumulative attribute importance score (97.50%) and drop the rest (which scored less than 1% each).

3 EXPERIMANTATION AND RESULTS

In this experiment, we used as input the 23 most important attributes according to ranking selection. In addition, accuracy was considered as metric of performance and the 10-fold cross validation procedure was used to evaluate each machine algorithm. We compared 5 different classifiers based on their accuracy: Naïve Bayes, SVMs, 5NN, Random Forests and a four layer fully connected Deep Neural Network (DNN) utilizing the dropout technique to reduce the data overfitting effect [10]. In Table 1 the performance results of the 5 classifiers are assessed. The DNN has the best accuracy with 71.50%. Moreover, the standard deviations of the results remain rather stable.

4 CONCLUSION

Many adults and older workers are spending more of their working hours sitting, which is a high-risk factor for cardiometabolic disease and all-cause mortality. In this paper we constructed a dataset regarding sedentary lifestyle and occupation, based on the ELSA observations. The assessment of the experimental results based on the accuracy of the 5 different classifiers demonstrates the superiority of the proposed DNN.

As future work, due to the sensitivity and complexity of healthcare data, we consider the examination of more complex DNN topologies and the incorporation of autoencoders [10] in order to improve the prediction accuracy and reduce the data noise.

Table 1: Classification Performance Comparison

Classifier	Accuracy (%)	Std. Deviation
Naive Bayes	68.88	1.35
SVMs	69.58	1.61
5NN	63.21	1.73
Random Forests	69.10	1.41
DNN	**71.50**	1.43

AKNOWLEDGMENTS

This work has been partially supported by the SmartWork project (GA 826343), EU H2020, SCI-DTH-03-2018 - Adaptive smart working and living environment supporting active and healthy ageing.

REFERENCES

[1] James Banks, GD Batty, K Coughlin, K Deepchand, M Marmot, James Nazroo, Z Oldfield, N Steel, Andrew Steptoe, M Wood, and P A Zaninotto. 2019. English Longitudinal Study of Ageing: Waves 0-8, 1998-2017. [data collection]. 29th Edition. UK Data Service. SN: 5050. DOI:https://doi.org/10.5255/UKDA-SN-5050-16

[2] Sally A M Fenton, Jet J C S Veldhuijzen van Zanten, Joan L Duda, George S Metsios, and George D Kitas. 2018. Sedentary behaviour in rheumatoid arthritis: definition, measurement and implications for health. *Rheumatology* 57, 2 (February 2018), 213–226. DOI:https://doi.org/10.1093/rheumatology/kex053

[3] Robin Genuer, Jean Michel Poggi, and Christine Tuleau-Malot. 2010. Variable selection using random forests. *Pattern Recognit. Lett.* (2010). DOI:https://doi.org/10.1016/j.patrec.2010.03.014

[4] Juliet A. Harvey, Sebastien F.M. Chastin, and Dawn A. Skelton. 2015. How Sedentary Are Older People? A Systematic Review of the Amount of Sedentary Behavior. *J. Aging Phys. Act.* 23, 3 (July 2015), 471–487. DOI:https://doi.org/10.1123/japa.2014-0164

[5] Joshua J Joseph, Justin B Echouffo-Tcheugui, Sherita H Golden, Haiying Chen, Nancy Swords Jenny, Mercedes R Carnethon, David Jacobs, Gregory L Burke, Dhananjay Vaidya, Pamela Ouyang, and Alain G Bertoni. 2016. Physical activity, sedentary behaviors and the incidence of type 2 diabetes mellitus: the Multi-Ethnic Study of Atherosclerosis (MESA). *BMJ Open Diabetes Res. Care* 4, 1 (June 2016), e000185. DOI:https://doi.org/10.1136/bmjdrc-2015-000185

[6] Otilia Kocsis, João Quintas, Hermie J. Hermens, Miriam Cabrita, Carina Dantas, Sonja Hansen, Tom Ziemke, Valentina Tageo, Paula Dougan, Konstantinos Moustakas, Nikos Fakotakis, Charalampos Vassiliou, Androniki Toska, Gregg C. Vanderheiden, Achilleas Stergiou, Dimitrios Amaxilatis, and André Pardal. 2019. SmartWork. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '19*, ACM Press, New York, New York, USA, 435–441. DOI:https://doi.org/10.1145/3316782.3322766

[7] Laust H. Mortensen, Ilene C. Siegler, John C. Barefoot, Morten Grønbæk, and Thorkild I.A. Sørensen. 2006. Prospective Associations between Sedentary Lifestyle and BMI in Midlife*. *Obesity* 14, 8 (August 2006), 1462–1471. DOI:https://doi.org/10.1038/oby.2006.166

[8] Sharon Parry and Leon Straker. 2013. The contribution of office work to sedentary behaviour associated risk. *BMC Public Health* 13, 1 (December 2013), 296. DOI:https://doi.org/10.1186/1471-2458-13-296

[9] Pekka Siirtola, Riitta Pyky, Riikka Ahola, Heli Koskimaki, Timo Jamsa, Raija Korpelainen, and Juha Roning. 2014. Detecting and profiling sedentary young men using machine learning algorithms. In *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, IEEE, 296–303. DOI:https://doi.org/10.1109/CIDM.2014.7008681

[10] Ian H. Witten, Eibe Frank, Mark A. Hall, and Christopher J. Pal. 2016. *Data Mining: Practical Machine Learning Tools and Techniques*. DOI:https://doi.org/10.1016/c2009-0-19715-5