Virtual Coach:
Predict Physical Activity Using a Machine Learning Approach

Talko Dijkhuis[1,2], Johan Blok[1] and Hugo Velthuijsen[1]

[1]Hanze University of Applied Sciences, Institute for Communication, Media & IT, Groningen, The Netherlands

[2]University of Groningen, Johann Bernoulli Institute for Mathematics and Computer Science, Groningen, The Netherlands

[1]{t.b.dijkhuis, j.blok, h.velthuijsen}@pl.hanze.nl

Abstract—One of the main causes of numerous health problems is a lack of physical activity. To promote a more active lifestyle, the Hanze University started a health promotion program. Participants were motivated to reach their daily goal of physical activity by means of an activity tracker in combination with two-weekly coaching sessions. Employing the data of the experiment, we investigated the manners in which the predictability of physical activity of a participant during the day can be improved. The collected step count data was used to construct personalised machine learning models, by taking into account the difference between physical activities during weekdays on the one hand and weekends on the other hand. The training of algorithms per participant in combination with the time-slices weekdays, weekend and the whole week improves the accuracy of the prediction model. The performance of the models improves even further when the individualised time-sliced models are combined. More contextual data, like free time and working hours, might even extend the accuracy. The use of personalised prediction models, based on machine learning and time slices, could become an addition in preventive personalized eHealth systems and mobile activity monitoring. For instance, this can constitute as a viable addition to a virtual coaching system to help the participants to reach their daily goal. As the individualised models allow for predictions of the progression of the physical activity during the day, they enable the virtual coaching system to intervene at the appropriate moment in time.

Keywords—preventive eHealth systems; monitoring physical activity; machine learning; prediction; virtual coach.

I. INTRODUCTION

An unhealthy lifestyle with insufficient daily physical activity shortens life expectancy. Not meeting the recommended level of physical activity is associated with 5.3 million deceases globally in 2008 [1]. Lack of physical activity is also associated with a decreased quality of life, lower levels of social participation, and disability to work. In the workplace employees with low and medium physical activity have a 2.4-3.5 fold higher rate of unplanned illness-related absenteeism compared to people who meet the Centers for Disease Control and Prevention (CDC) guideline of 150 min/week [2].

The negative effects of lacking physical activity have fostered a novel initiative at the Dutch University, Hanze University of Applied Sciences (HUAS). The university started an initiative to promote a healthy lifestyle and physical activity during the workday called (in Dutch): Het Nieuwe Gezonde Werken (The New Healthy Way of Working; HNGW). This initiative on promotion of a healthy lifestyle included a focus on the improvement of physical activity. Participants got an activity tracker to increase the awareness of their daily progress in achieving their goals in terms of numbers of steps. The daily feedback of the activity tracker was complemented with a fortnightly coaching session on the lifestyle and the physical activity. However, the feedback of the activity tracker and its platform didn't provide the participant with timely personalised feedback. Neither was the coach timely informed with information on the participant to enable a personalised intervention. Furthermore, current activity trackers do not provide a probability of reaching the daily goal or take the difference between weekdays and weekend into account, although this difference is known for a different level of activity [3].

In this paper, we propose a personalised, flexible machine learning based model that enables personalized eHealth being supported by preventive systems on activity tracking. The personalised model enables feedback on a participant's probability of reaching his or her daily activity goal. The first section introduces the state of the art on measuring activity levels, the use of machine leaning and monitoring. Subsequently, we describe the study on health promotion at HUAS, the collected dataset on daily physical activity of the participants, the method of statistical analysis of the results trained algorithms, and the selection and training of the algorithms. In the third section, we present the results of the training of the algorithms and the statistical analysis. The conclusion on the results and a short discussion on future work finish this paper.

II. STATE OF THE ART

Activity trackers provide a measure for the number of steps humans make and enable monitoring. Adding a step counter to physical therapy or counselling was effective in some groups [4] [5]. The collection of step data is not only effective for therapy or counselling, it is also an intervention mechanism in itself [6]. Only the fact of using an activity tracker could motivate physical activity and improvement of health [7]. To improve on physical activity in combination with activity tracking monitoring, coaching is helpful. Perceiving the information personal and in context and timely is important for the effectiveness of (e)Coaching [8]. The participant needs to receive the information and the advice while it is relevant. To the best of our knowledge no studies exist on the use of activity trackers in combination with machine learning

algorithms to establish *individualized* models or studies on *individualized* models used in preventive systems on monitoring activity helping the participant to improve his or her physical behaviour.

III. METHODS

In this section, we present the study design of the HNGW, the data set we used to train the algorithms and the methods used for statistical analysis.

A. Study design

The study data stems from the HNGW project. Forty-eight healthy employees were recruited from the HUAS. The 48 participants were divided according to age, gender, BMI, and baseline self-reported health prior to being randomized into two groups. Group A followed a twelve-week health promotion intervention; the other group, group B, served as a control first and thereafter received the intervention. The outcome measures included, among other values, the daily steps. The daily steps were measured with the Fitbit Flex, which is known to be a trustworthy and valid activity tracker for step count and suitable for health promotion programs [9].

1) Data set: In order to prepare the available minute step data as input for training the algorithms, we followed a step-by-step approach. First, we performed a data pre-processing step to remove the incomplete records from the data set. We also eliminated all records per day whenever no step was gathered during that day. Second, we constructed an hourly summarised data set with several new derived variables representing:

1) the day of the week (range 0 - 6)
2) hour of day (range 0 - 23)
3) week of the year (range 0-52)
4) year (2014-2015)
5) a cumulative sum of the steps per hour

Third, a workday is defined as the weekdays Monday till Friday. The normal working hours at the university are between 08:00AM and 05:00PM. The project tried to motivate the participants to walk at least a part of the distance they daily commute. As a consequence, the hours of interest are the combination of the working hours and commuting. Therefore, we sliced the dataset such that it only contains the number of steps per hour, per workday between 07:00AM and 06:00PM. Fourth, a weekend is defined as Saturday and Sunday. In order to enable comparison with the weekdays, we sliced the data set for the weekend in the same way as the workdays, per weekend day between 07:00AM and 06:00PM. Fifth, partial sum of steps per hour throughout the day was included. Sixth, a column was added comprising the average number of steps at 06:00PM calculated over all weeks. For this average the amount of steps between 7:00AM and 6:00PM was considered. This column was regarded as a threshold in order to determine the outcome column. Finally, we constructed a binary outcome variable based on the threshold.

2) Statistical Analysis: Four different algorithms were trained. To compare the performance of the algorithms, we used the confusion matrix method to classify the difference between the predicted value and the actual value. A confusion matrix provides an overview of the true positives (TP; a predicted a 'true' and the actual data contained a 'true'), true negatives (TN; the model predicted a 'false' and the actual data was a 'false'), false positives (FP; the model predicted a 'true' label, but the actual data was a 'false'), and false negatives (FN; the model predicted a 'false' label, but the data was 'true') of a model. The confusion matrix served as a basis for the calculation of the performance measure F1-score [10].

The F1-score was calculated for each model, the F1-score has a range of zero to one, one is the best score. To calculate the F1-score, two other metrics known as the precision and the recall are used. Precision is the proportion of the true positives and the false negatives, and is calculated as $\frac{TP}{(TP+FN)}$. Recall is the true positive rate, which is calculated as $\frac{TP}{(TP+FP)}$. Using precision and recall, the F1-score is calculated as $2 \cdot \frac{Precision \cdot Recall}{Precision+Recall}$.

B. Selection of algorithms

The goal is to predict, during the day, whether a participant will reach his daily number of steps. This is known as a classification problem. The selection of the best algorithm is a matter of trial and error. It is generally agreed upon that it's not possible to determine the best performing algorithm upfront [11]. The general approach for solving this problem is very similar to the travelling salesman problem [12]. Although there are classes of algorithms which are more suitable for different types of problems. One of the biggest open-source community on machine learning, scikit-learn.org provides a 'flowchart' with rough indications which algorithms may perform best [13]. We choose four possibly well-performing algorithms: (i) ADAboost (ADA), (ii) Decision Tree (DT), (iii) Random Forest (RF), and (iv) Stochastic Gradient Descent (SGD). After splitting the data in a training- and a test set, the performance of each of the algorithms was calculated.

Firstly, based on the whole training set the F1-score of each algorithm was determined. Secondly, the algorithms were trained on the individualized training data utilizing three different time slices of the dataset and the trained algorithms were converted into to individualized time slice based models (TSM):

TSM1:work week (Monday-Friday)
TSM2:weekend (Saturday, Sunday)
TSM3:whole week (Monday-Sunday)

The result of the training was of 12 different models per participant (TSM 1-3 times the four algorithms). Next the ranking and the overall best performing algorithm was determined.

Thirdly, for the three personalized time sliced models of the overall best performing algorithm, the F1-score was calculated using the complete data set. TSM 1 was used to calculate the

F1-score for the workweek, TSM 2 was used to calculate the F1-score for the weekend, and TSM 3 was used to calculate two F1-scores, respectively for the work week and for the weekend.

Fourthly, the combination of week F1-score and weekend F1-score of the diverse time slice models were studied on the performance.

IV. RESULTS

The group F1-scores per algorithm were for Random Forest 0.89, Decision Tree 0.88, ADAboost 0.69, and Stochastic Gradient Descent 0.44. Application of the individualized component and time slices slightly improved the performance. Only ADAboost showed big differences on the F1-score. Figure 1 displays the results of the average of the individual scores on the subsequent algorithms and time-slices.
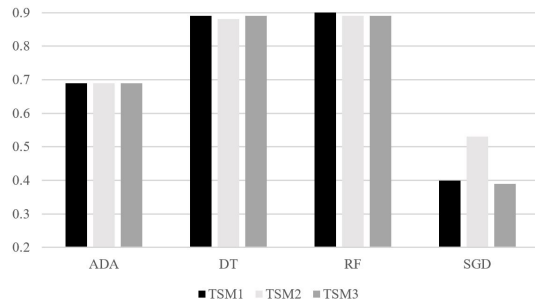

Fig. 1. Average F1-score of the time-sliced models over all participants.

Table I represents the numbers of the average of the individual scores.

TABLE I. AVERAGE F1-SCORE OVER ALL PARTICIPANTS OF THE TIME-SLICED MODELS.

	ADA	DT	RF	SDG
TSM1	0.69	0.89	0.9	0.4
TSM2	0.69	0.88	0.89	0.53
TSM3	0.69	0.89	0.89	0.39

On a group level the best performing algorithm is Random Forest. The Random Forest based, individual time sliced models in different combinations resulted in diverse best combinations. Table II states the diverse combinations of the time slice models and the F1-score. For 29 participants there is one ideal combination, for 12 participants there are 2 best combinations and for 3 participants the diverse combinations of the time slice models perform equally. All combinations of different individualised time slice models outperformed the group result stated in Table II.

V. CONCLUSION

The individualisation of the machine learning models improved the F1-score in comparison to the group level F1-score. The best performing algorithm was the Random Forest algorithm. Application of the literature based thesis concerning the difference of physical activity between week days and

TABLE II. COMBINATIONS OF TIME-SLICES AND THEIR F1-SCORE.

	A & B	A & D	C & B	C & D
one best	6	11	9	3
F1-score	0.95	0.95	0.92	0.93
standard deviation	0.03	0.02	0.03	0.02
two equally best	6	5	7	6
F1-score	0.95	0.95	0.95	0.95
standard deviation	0.03	0.02	0.03	0.02
all equal	3	3	3	3
F1-score	0.96	0.96	0.96	0.96
standard deviation	0.02	0.02	0.02	0.02

A: TSM1, work week (range: Monday-Friday)
B: TSM2, weekend (range: Saturday, Sunday)
C: TSM3, work week (range: Monday-Friday)
D: TSM3, weekend (range: Saturday, Sunday)

weekend for training different algorithms improved the F1-score. It is recommended to construct time sliced weekend and week models per individual and calculate which combination of models performs best. To improve the performance of the individualised models in the future, contextual data that influences physical activity, like free time, regular physical activity, and illness, may be taken into account. The individualisation of the predictive models enables automated personalised timely coaching. The results of this paper will be applied in the preventive eHealth virtual coach platform as suggested by Blok et al. [14]. A possible future direction is to create a model per day per individual.

ACKNOWLEDGMENT

REFERENCES

[1] I. Min-Lee *et al.*, "Effect of physical inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and life expectancy," *The Lancet*, vol. 380, no. 9838, pp. 219–229, 2012. [Online]. Available: http://dx.doi.org/10.1016/S0140-6736(12)61031-9

[2] E. Losina, H. Y. Yang, B. R. Deshpande, J. N. Katz, and J. E. Collins, "Physical activity and unplanned illness-related work absenteeism: Data from an employee wellness program," *PLoS ONE*, vol. 12, no. 5, pp. 1–8, 2017.

[3] C. E. Matthews *et al.*, "Sources of Variance in Daily Physical Activity Levels in the Seasonal Variation of Blood Cholesterol Study," *American Journal of Epidemiology*, vol. 153, no. 10, pp. 987–995, 2001.

[4] H. J. de Vries, T. J. Kooiman, M. W. van Ittersum, M. van Brussel, and M. de Groot, "Do activity monitors increase physical activity in adults with overweight or obesity? A systematic review and meta-analysis," *Obesity*, vol. 24, no. 10, pp. 2078–2091, oct 2016. [Online]. Available: http://doi.wiley.com/10.1002/oby.21619

[5] M. Miyauchi *et al.*, "Exercise Therapy for Management of Type 2 Diabetes Mellitus: Superior Efficacy of Activity Monitors over Pedometers," *Journal of Diabetes Research*, vol. 2016, pp. 1–7, sep 2016. [Online]. Available: https://www.hindawi.com/journals/jdr/2016/5043964/

[6] L. A. Cadmus-Bertram, B. H. Marcus, R. E. Patterson, B. A. Parker, and B. L. Morey, "Randomized Trial of a Fitbit-Based Physical Activity Intervention for Women." *American journal of preventive medicine*, vol. 49, no. 3, pp. 414–8, sep 2015. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/26071863http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4993151

[7] Z. H. Lewis, E. J. Lyons, J. M. Jarvis, and J. Baillargeon, "Using an electronic activity monitor system as an intervention modality: A systematic review," *BMC Public Health*, vol. 15, no. 1, p. 585, dec 2015. [Online]. Available: http://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-015-1947-3

[8] M. Gerdes, S. Martinez, and D. Tjondronegoro, "Conceptualization of a Personalized eCoach for Wellness Promotion," in *Proceedings of ACM 11th EAI International Conference on Pervasive Computing*. Barcelona: Association for Computing Machinery, 2017.

[9] J. Wang, R. Chen, X. Sun, M. F. H. She, and Y. Wu, "Recognizing human daily activities from accelerometer signal," *Procedia Engineering*, vol. 15, pp. 1780–1786, 2011, URL: http://dx.doi.org/10.1016/j.proeng.2011.08.331 [accessed:2018-01-31].

[10] S. Schoeppe *et al.*, "Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 13, no. 1, p. 127, dec 2016, URL: http://ijbnpa.biomedcentral.com/articles/10.1186/s12966-016-0454-y [accessed:2018-01-31].

[11] D. H. Wolpert, "The lack of a priori distinctions between learning algorithms," *Neural Computation*, vol. 8, no. 7, pp. 1341–1390, 1996, URL: http://www.mitpressjournals.org/doi/10.1162/neco.1996.8.7.1341 [accessed:2018-01-31].

[12] S. Raschka and V. Mirjalili, *Python Machine Learning*. Packt Publishing Ltd, 2015.

[13] "scikit learn,choosing the right estimator," URL: http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html [accessed:2018-01-31].

[14] J. Blok, A. Dol, and T. Dijkhuis, "Toward a Generic Personalized Virtual Coach for Self-management : a Proposal for an Architecture," no. c, pp. 105–108, 2017.