German Medical Data Sciences 2023 — Science. Close to People.
R. Röhrig et al. (Eds.)

doi:10.3233/SHTI230711

Challenge of Detecting Personal Deviations and Trends in Sensor Based Activity Data

Miriam LINGG[a,b,1], Chantal BEUTTER[a], Stefan SIGLE[a], Daniel ZSEBEDITS[a], and Christian FEGELER[a,b]
[a] *MOLIT Institute gGmbH, Heilbronn, Germany*
[b] *Heilbronn University of Applied Sciences, Heilbronn, Germany*

Abstract. INTRODUCTION Physical activity and health are closely linked. Therefore, monitoring movement behavior is of great interest e.g., to monitor a patient's physical state. Nowadays it is easy to record movement with a smartphone. The aim of this work was to develop a concept to detect trends based on personalized movement behavior recorded with a smartphone. **METHODS** A first prototype with a control chart was designed. Since this approach did not prove suitable for analyzing activity data for trends in practice, a second prototype was subsequently developed with a statistical trend test (Mann-Kendall test (MK test)). It was extended by the Yue-Wang correction approach to be able to deal effectively with serial correlation. Furthermore, the traditional trend modeling using Theil-Sen slope was extended by three additional models to be able to represent non-linear trend shapes. **RESULTS** Movement behavior can be highly variable, which leads to wide control limits when using control charts. As the lower control limit was always in the negative range the use of a control chart was impossible for this use case. The evaluation results of the second prototype confirm the choice of a non-parametric test, as well as the decision for the Yue-Wang correction factor. Furthermore, it could be determined that the MK test is robust against outliers. The number of detected trends increases with increasing significance level. The MK test is also suitable for detecting step-like trends. **CONCLUSION** Live trend detection is not straightforward with the MK test but can be simulated by overlapping time periods. In the future, trend modeling should be extended even further, as it plays a major role in the concept. The sensitivity of the test can be increased by means of various parameters.

Keywords. Exercise, Smartphone, Trend Detection, Electronic Data Processing, Statistics, Nonparametric

1. Introduction

Physical activity and health are closely related in many aspects. For example, regular exercise has been shown to reduce the risk of developing depression and has a positive impact for people who are already suffering from depression [1]. Existing symptoms can be reduced by exercise, making it a good complement to conventional drug and psychotherapeutic treatment [2]. Apart from the research field of depression, exercise is also an important aspect in the field of oncology and cardiovascular diseases as it may lead to a lower risk for contradicting the disease and reduced mortality [3]. But it is not

[1] Corresponding Author, Miriam Lingg, MOLIT Institut gGmbH, Im Zukunftspark 10, 74076 Heilbronn, Germany; E-mail: miriam.lingg@molit-service.de.

only important to track overall movement, but also to detect changes in the activity behavior. In cancer patients for example, a decrease in physical activity may indicate new-onset problems, both physical and psychological. There is also preliminary evidence that a decrease in physical activity during active cancer treatment may indicate acute toxicity such as renal insufficiency [4].

In conclusion, physical activity has a major impact on health and the progression of disease. Therefore, monitoring physical activity behavior for changes is of great interest. Nowadays, people's physical activity behavior can be easily recorded in everyday life. For example, smartphones can be used for this purpose, which have integrated sensors that can be used to detect movement [4]. The goal of this work was to develop a concept to examine movement behavior recorded by a smartphone detecting changes and trends longitudinally.

2. Methods

For evaluation subjects consent to track their movement for a fixed amount of time. To easily record the movement in everyday life, a smartphone app is used. However, the app does not record the user's movements itself, but derives active move minutes recorded by Google Fit for Android users and Apple Health for iOS users. To create an option for activities not detected by those two health apps, such as swimming, the app offers the option of manually entering activity and movement data [5]. To test our approach, two healthy female volunteers produced a total of three test series. Test person 1 provided two test series, whose courses can be seen in Figure 1. The first test series covered 51 days, the second test series 119 days. Test person 2 generated a third test series over 119 days, the course of which can be seen in Figure 2.

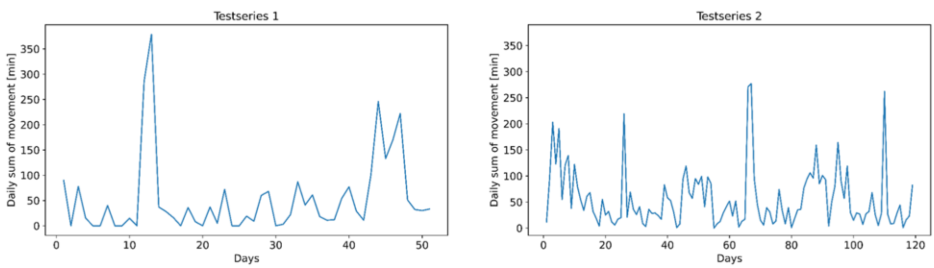

Figure 1. Course of daily movement in minutes of test person 1.

The test series were split into overlapping periods of 14, 21 and 28 days. These periods started every 7 days and thus overlapped by 7, 14 and 21 days. The 51-day test series was divided into 6 periods of 14 days, 5 periods of 21 days and 4 periods of 28 days. The two 119-day time series were split into 16 periods of 14 days, 15 periods of 21 days, and 14 periods of 28 days. In addition, the test persons indicated for each day whether it was "normal" or not. "Normal" days are those on which the user was not ill, injured or on holiday. The movement data of the test persons was extracted in JSON format. The sum of the movement minutes is then calculated for each day, and these daily totals should be examined for trends and deviations.

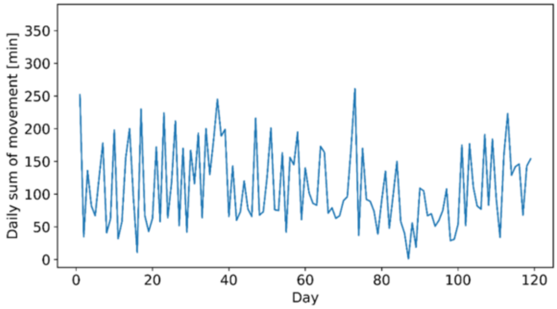

Figure 2. Course of daily movement in minutes of test person 2.

There are different ways to detect trends in time series. On the one hand there are control charts, on the other hand there are statistical trend tests [6]. In theory, the control charts are even more suitable than the trend tests, because they are better suited to check continuous values for anomalies. A statistical trend test, on the other hand, evaluates whether there is a trend in a fixed period. This means that a new test must always be applied when new data is available.

Therefore, a prototype with an Shewhart-control chart was designed first. The basic procedure for a control chart is as follows: first, baseline data is collected, which is then used to calculate the warning and intervention limits. Afterwards, new values can simply be entered into the control chart, and it is immediately apparent whether the new value is conspicuous or not [6].

Within the research a second prototype was subsequently developed with a statistical trend test. For this purpose, the MK test was used. To perform it, several variables must first be calculated. The following formulas are taken from the work of Yue and Wang [7] unless otherwise indicated. First, the test statistic S is calculated according to the following formula:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn\left(X_j - X_i\right) \tag{1}$$

where n describes the length of the data set, and X_j, X_i are the values of the data series at are the locations j, i. The signum function $sgn(\theta)$ is calculated as follows.

$$sgn(\theta) = \begin{cases} 1 & if\ \theta > 0 \\ 0 & if\ \theta = 0 \\ -1 & if\ \theta < 0 \end{cases} \tag{2}$$

Then, the variance of the test statistic V(S) is calculated as described in equation (3), where n describes the length of the data set. The back part of the equation represents an adjustment of the variance for the case of tied values. Ties occur when the values in the time series are identical at several different points in time. The extent of a tie is described by how often a value is present in the time series. t_i is the number of ties of extent i.

$$V(S) = \frac{n(n-1)(2n+5)-\sum_{i=1}^{n} t_i i(i-1)(2i+5)}{18} \tag{3}$$

To effectively address serial correlation, the MK test is extended by the Yue-Wang correction factor CF. To calculate it, the trend must first be modeled and removed from the original time series. Often the Theil-Sen slope is used for this purpose. It is a non-parametric estimator of a linear trend [7]. To be able to estimate other, non-linear trend forms, the trend modeling was extended by three additional models: the square root model, the logarithmic model, and the multiplicative model. First, the parameters of the models are calculated. Then, the coefficient of determination R^2 is calculated for all four models, which is a measure of how well a model fits the underlying data. The model with the highest R^2 value is selected and the trend estimated with this model is removed from the time series [8,8]. Then, the correction factor is calculated using the following equations and the modified variance V*(S) is finally obtained.

$$V^*(S) = V(S) \cdot 1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right) \cdot \rho_k \tag{4}$$

Where ρ_k is the lag-k autocorrelation coefficient, which can be represented by the lag-k correlation coefficient r_k.

$$r_k = \frac{\frac{1}{n-k}\sum_{t=1}^{n-k}(X_t - \bar{X}_t)(X_{t+k} - \bar{X}_t)}{\frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X}_t)^2} \tag{5}$$

where n is the number of data points, k is the lag of the autocorrelation coefficient and $\overline{x_t}$ is the mean value of the data according to the following equation.

$$\bar{X}_t = \frac{1}{n}\sum_{t=1}^{n} X_t \tag{6}$$

Finally, the standardized test statistic Z is calculated using the test statistic S, as well as the modified variance V*(S).

$$Z = \begin{cases} \frac{S-1}{\sqrt{V^*(S)}} & if\ S < 0 \\ 0 & if\ S = 0 \\ \frac{S+1}{\sqrt{V^*(S)}} & if\ S < 0 \end{cases} \tag{7}$$

The value of Z can then be directly compared to the corresponding quantile of the standard normal distribution. If the absolute value of Z exceeds the value of the quantile, the null hypothesis (no trend) is rejected at the selected significance level α. A negative value of Z indicates a falling trend, a positive value an increasing trend [7].

The second prototype was then implemented and evaluated. The self-implementation of the MK-test component was validated against reference implementations in R and Python. With the MK test, a non-parametric test was selected.

Therefore, it was first reviewed whether the generated test data in periods of two to four weeks are rather non-normally distributed to be able to evaluate the decision for this type of trend test. For this purpose, two statistical tests, Shapiro-Wilk test and Anderson-Darling test, with significance level $\alpha = 0, 05$ were applied. If at least one of the two tests concludes that a period is non-normally distributed, this period is to be considered as non-normally distributed. With the Yue-Wang expansion, an approach against serial correlation was selected, which is also effective against serial correlation in higher lags than 1. Therefore the significance of the autocorrelation coefficients was tested at a significance level of $\alpha = 0, 05$. Subsequently, evaluations were carried out to investigate the influence of outliers. A value was considered an outlier if it was 1.5 interquartile distances below the first quartile or 1.5 interquartile distances above the third quartile. To investigate the influence of the significance level on the test result, all time periods were tested at the significance levels 0.01, 0.05 and 0.1 Finally, an experiment was carried out to check whether the MK test is also capable of detecting step-like trends.

3. Results

3.1. First prototype with control chart

The normal movement behavior of people can vary greatly. To calculate a baseline as reference, the first 21 "normal" days were used. Non-normal days were excluded. However, when calculating the limits, the lower negative intervention limit was in the negative range for all three test series. An example control chart of test person 2 can be seen in Figure 3. Also, changing the baseline length to 14 or 28 days did not result in a positive threshold. After that, another attempt was made to collect the baseline in a different way. In a 21-day period, at least 16 normal days must be present to calculate the limits from the corresponding values. With this alternative, too, all lower control limits were in the negative range. This result showed that a normal movement behavior leads to a high standard deviation and consequently to very wide warning and control limits. In this use case, an intervention limit in the negative range would mean that an anomaly would only be detected if there was a negative number of minutes of movement for a day, which is not possible. Since the detection of anomalies and trends is based directly on those limits, it is not possible to observe them with a control chart.

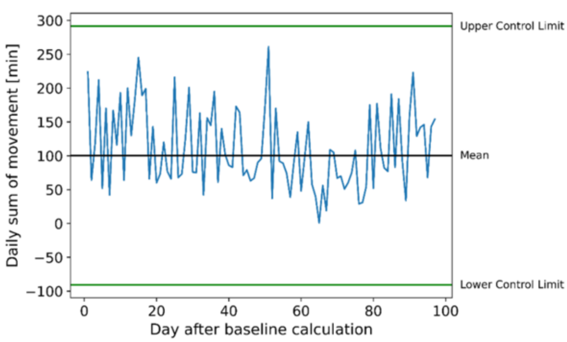

Figure 3. Control chart of test person 2. Baseline and corresponding control limits were calculated with the values of the first 21 ‘normal’ days.

3.2. Implementation of the second prototype with the MK test

The concept of the second prototype was fully self-implemented using the Java programming language. The developed component can be easily integrated as a .jar file as a library into further projects, as Maven was used as a build tool. The access is provided via a fixed interface. The desired significance level of the test and Fast Healthcare Interoperability Resource (FHIR) observations are entered as input parameters. Individual movement entries with start time and duration are stored in the FHIR observations. As output, the user receives information on the best-fitting model, as well as information on whether there is a trend for the period in question. The implementation consists of three main components. The first component is used to pre-process the movement data and to calculate the total amount of activity per day. The second component performs the trend modelling and calculates which trend model best fits the underlying data. In the last component, the actual MK test is implemented and determines whether a trend is present or not in the given period.

3.3. Results of evaluation of the second prototype

When evaluating the test data for normal distribution, it was found that the longer the time periods considered, the more likely they were not normally distributed, which can be seen in Table 1. For the 14-day periods, approximately 47 % (18/38 periods) were not normally distributed. For 21-day periods, the proportion of non-normally distributed periods increased to approximately 66 % (23/35 periods) and for 28-day periods to approximately 81 % (26/32 periods). These results confirm the choice of a non-parametric test.

Table 1. Number of normally and non-normally distributed periods at different period lengths.

	14 Days	21 Days	28 Days
Normally distributed	20	12	6
Not normally distributed	18	23	26

When evaluating the autocorrelation of the test data, significant autocorrelation was found in a total of 22 time periods. In about 32 % of the cases (7/22 periods), the autocorrelation was at a higher lag than 1. These results confirm the choice of the Yue-Wang correction approach, which in contrast to other alternatives can also consider serial correlation in higher lags than 1.

Furthermore, it could be determined that the MK test is not influenced by outliers. Upward outliers at the beginning of an otherwise apparently trend-free time series do not lead to the MK-test indicating a falling trend. Upward outliers at the end of an otherwise apparently trend-free time series do not lead to the MK-test indicating a positive trend. Also, upward outliers at the beginning of an otherwise apparently rising time series do not prevent the detection of a rising trend. Finally, upward outliers at the end of an otherwise apparently falling time series do not prevent a falling trend from being detected.

The evaluation of the influence of the significance level on the number of detected trends showed that the more trends are detected, the higher the significance level is chosen, which is shown in Table 2. Changing the significance level from $\alpha = 0.01$ to $\alpha = 0.05$ and $\alpha = 0.1$ resulted in an increase in the number of detected trends from 12 to 17 and to 18 for the 14-day periods, and from 15 to 17 and then to 19 for the 21-day periods. For the 28-day periods, the number increased from 13 to 17 and to 19 detected trends. Thus, a higher power of the MK test can be achieved with a larger chosen significance level.

Table 2. Number of detected trends depending on the selected significance level α for different period lengths.

	14 Days	21 Days	28 Days
$\alpha = 0.01$	12	15	13
$\alpha = 0.05$	17	17	17
$\alpha = 0.10$	18	19	19

In the experiment on step-like trends, subject 1 obtained the expected results in all 12 periods concerned. For subject 2, the expected results were met in 9 out of 12 affected time periods. These results suggest that the MK test is also capable of registering step-like changes in movement behavior as trends.

4. Discussion

To examine time series for trends, a first prototype was developed using the Shewhart control chart. It is best suited in theory but has not proven itself in practice. Motion data is too variable, which leads to the warning and control limits becoming too wide. However, it should be noted that in the event of modified use cases, the control charts should be considered again. For example, it would be conceivable in the future that users could set targets for their daily or weekly exercise in the app. This target time could then be stored in a control chart as a targeted target value. A second prototype was then developed with the MK test. The test is designed to examine data of fixed time periods. Live trend detection is not so simply possible with it since a new test must be carried out for new values. To imitate a live detection, tests were started every 7 days staggered over a fixed period of time. However, when interpreting the results of successive tests, the *multiple test problem* must be considered. For example, if there are increasing trends at the 0.05 significance level in two successive 14-day periods, this could quickly lead to the assumption that there is a 21-day trend because the two periods overlap by 7 days. But it must be considered that this assumption is no longer made with significance level 0.05, since the probability of committing at least one type I error increases with the number of tests performed [9].

In the developed concept, the Yue-Wang correction factor is used to effectively address serial correlation. For this, the trend must first be estimated and then removed. In addition to the classical modeling with Theil-Sen Slope, three further models were used, which can also represent non-linear trends. An extension of the trend modeling for further non-linear trend forms should be considered in the future. In approximately 27 % of the analyzed time periods, the best-fitting model was not the linear Theil-Sen slope. However, an accurate estimate of the trend is important to the procedure. The better the trend can be estimated and removed the better a possible serial correlation can be addressed.

To increase the sensitivity of the test, several things can be modified. First, the size of the selected significance level. The larger this is chosen, the greater the power of the test. Therefore, the significance level should be chosen with a value of 0.1. The power further increases with increasing length of the time series. However, it must be noted that a longer period time leads to an increased latency for the detection of a trend and thus the detection of possibly deteriorations in the health status, as the trend detection is not performed until the period is completed. Also, the test problem was formulated as two-sided to be able to detect both rising and falling trends. A further increase of the power could be achieved if the test problem was formulated exclusively one-sided. This would mean, for example, that the test would be limited to the detection of falling trends only. Two further parameters that influence the sensitivity are the trend and the variance in the data. The stronger the trend and the lower the variance, the greater the power of the MK test. However, these cannot be influenced in practice, as they depend exclusively on the user's movement behavior.

5. Conclusion

The goal of this work was to develop a concept that makes it possible to examine the movement behavior on trends. After the implementation of a first prototype with a Shewhart control chart, it was found that control charts are not suitable for the given use case. A second prototype was then developed using the MK test. With the MK-test a non-parametric trend test was selected, which is robust against outliers and capable able to detect step-like trends. This was extended to include the Yue-Wang correction approach to effectively address serial correlation. In addition, an extension of the trend modeling was also carried out, which makes it possible to ideally represent non-linear trends as well. This proved valuable during the evaluation, as trends in time series are often non-linear. For live trend detection via overlapping time periods, the multiple testing problem must be considered when interpreting the results. The sensitivity of the test can be increased by choosing a larger level of significance as well as using longer time series. A further increase can be achieved by a reformulation to a one-sided test problem. The designed concept was then implemented and finally evaluated. The test subjects were healthy, but it can be assumed that the results can also be applied to sick people. With the implemented component, a system was developed that can be easily integrated into other projects as a library.

Declarations

Conflict of Interest: The MOLIT Institute is a non-profit organization, funded by donation. The last author is one of the founders of the MOLIT Institute.

Availability of data: The data and code that support the findings of this study are available from the corresponding author, ML, upon reasonable request.

Author contributions: ML: conception, implementation, and data analysis of the work, writing of the manuscript; ML, CB: data acquisition; CB, SS, DZ: revising progress of work, revising and editing of the manuscript; CF: substantial revising of the manuscript, supervising professor. All authors approved the manuscript in the submitted version and take responsibility for the scientific integrity of the work.

References

[1] F.B. Schuch, D. Vancampfort, J. Firth, S. Rosenbaum, P.B. Ward, E.S. Silva, M. Hallgren, A. Ponce De Leon, A.L. Dunn, A.C. Deslandes, M.P. Fleck, A.F. Carvalho, and B. Stubbs, Physical Activity and Incident Depression: A Meta-Analysis of Prospective Cohort Studies. Am J Psychiatry **175** (2018), 631–648.
[2] A. Kandola, G. Ashdown-Franks, J. Hendrikse, C.M. Sabiston, and B. Stubbs, Physical activity and depression: Towards understanding the antidepressant mechanisms of physical activity. Neuroscience & Biobehavioral Reviews **107** (2019), 525–539.
[3] A. McTiernan, C.M. Friedenreich, P.T. Katzmarzyk, K.E. Powell, R. Macko, D. Buchner, L.S. Pescatello, B. Bloodgood, B. Tennant, A. Vaux-Bjerke, S.M. George, R.P. Troiano, and K.L. Piercy, Physical Activity in Cancer Prevention and Survival: A Systematic Review. Med Sci Sports Exerc **51** (2019), 1252–1261.
[4] H.S.L. Jim, A.I. Hoogland, N.C. Brownstein, A. Barata, A.P. Dicker, H. Knoop, B.D. Gonzalez, R. Perkins, D. Rollison, S.M. Gilbert, R. Nanda, A. Berglund, R. Mitchell, and P.A.S. Johnstone, Innovations in research and clinical care using patient-generated health data. CA Cancer J Clin **70** (2020), 182–199.
[5] Lebensqualitätsstudie – Lion-App, https://lion-app.de/evaluation/ [cited 2023 March 27].
[6] P.U.B. Vogel, *Trending in der Pharmazeutischen Industrie,* Springer Fachmedien Wiesbaden GmbH, Wiesbaden, 2020.
[7] S. Yue and C. Wang, The Mann-Kendall Test Modified by Effective Sample Size to Detect Trend in Serially Correlated Hydrological Series. Water Resources Management **18** (2004), 201–218.
[8] K. Backhaus, B. Erichson, W. Plinke, and R. Weiber, *Multivariate Analysemethoden,* Springer Berlin Heidelberg, Berlin, Heidelberg, 2016.
[9] L. Fahrmeir, C. Heumann, R. Künstler, I. Pigeot, and G. Tutz, *Statistik,* Springer Berlin Heidelberg, Berlin, Heidelberg, 2016. p. 395