D4.4: Methods for inferring long-term behaviours from short-term behaviour information

Dissemination level: Public
Document type: Report
Version: 1.0.0
Date: February 27, 2019

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement #769553. This result only reflects the author's view and the EU is not responsible for any use that may be made of the information it contains.

Document Details

Project Number	769553
Project title	Council of Coaches
Title of deliverable	Methods for inferring short-term behaviour from multimodal sensor data
Due date of deliverable	February 28, 2019 (M18)
Work package	WP4
Author(s)	Oresti Banos (CMC), Kostas Konsolakis (CMC)
Reviewer(s)	Reshma Bangalore Kantharaju (SU)
Approved by	Coordinator
Dissemination level	Public
Document type	Report
Total number of pages	28

Partners

- University of Twente – Centre for Monitoring and Coaching (CMC)
- Roessingh Research and Development (RRD)
- Danish Board of Technology Foundation (DBT)
- Sorbonne University (SU)
- University of Dundee (UDun)
- Universitat Politècnica de València, Grupo SABIEN (UPV)
- Innovation Sprint (iSPRINT)

Abstract

This deliverable (D4.4) represents the contribution of Work Package 4 (WP4) to the inference of long-term behaviours from short-term behaviour data. Long-term behaviours are investigated, as part of task T4.2, in order to derive the user lifestyle, health and well-being states for the generation of dialog and argumentation content, which will be used in WP3 and WP5. Thus, this deliverable aims to research and develop the techniques for intelligently combining the behaviour primitives generated in T4.1 into more descriptive representations of behaviour.

Table of Contents

List of figures

List of tables

Symbols, abbreviations and acronyms

AR	Autoregression
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BJ	Box-Jenkins
CMC	Centre for Monitoring and Coaching
COUCH	Council of Coaches
D	Deliverable
EC	European Commission
GCM	Growth Curve Modelling
GMM	Growth Mixture Modelling
GPS	Global Positioning System
HBAF	Holistic Behaviour Analysis Framework
HCI	Human-Computer Interaction
HWES	Holt Winter's Exponential Smoothing
LCGA	Latent Class Growth Analysis
LGCM	Latent Growth Curve Modelling
M	Month
MA	Moving Average
MS	Milestone
RRD	Roessingh Research and Development
SARIMA	Seasonal Autoregressive Integrated Moving Average
SARIMAX	Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors
SEM	Structural Equation Modelling
SES	Simple Exponential Smoothing
STD	Standard Deviation
UPMC	Université Pierre et Marie Curie, Paris 6
UPV	Universitat Politècnica de València
UT	University of Twente
VAR	Vector Autoregression
VARMA	Vector Autoregression Moving-Average
VARMAX	Vector Autoregression Moving-Average with Exogeneous Regressors
WP	Work Package

1 Introduction

After presenting the initial concept of T4.1 on inferring short-term behaviours from sensor data, thoroughly elaborated in D4.2 and D4.3, this deliverable focuses on the inference of long-term behaviours from short-term behaviours. Long-term behaviours are investigated, as part of the task T4.2, in order to derive the user lifestyle, and by extension, get some insights into their health and well-being states. The long-term behaviour information will be further used by WP3 and WP5 for the generation of dialog and argumentation contents. In addition, studying long-term behaviours will enable the analysis of user behaviour patterns or trajectories, which will be further mined over time in T4.3 to identify relevant behavioural changes demanding a prompt or preventive intervention from a specific coach or the complete council. It is worth mentioning that the detected changes will be used in future work to trigger specific coaching actions as defined in T3.2 and executed by the Dialogue and Argumentation Framework in WP5.

The methods used for detecting long-term behaviours (a.k.a. routines), including the terminology and the techniques for behaviour analysis, are thoroughly investigated and presented in Section 3. A preliminary evaluation of these methods is performed for physical routines, as these have a chief impact on the use cases considered in the Council of Coaches project. This evaluation is shown in Section 4, while Section 5 defines the main outcomes of this deliverable.

2 Objectives

The main objective of this deliverable (D4.4) is to describe the methods developed for the inference of long-term behaviours or routines based on the aggregation of short-term behaviour data. Accordingly, this document aims to investigate and elaborate on the data processing and statistical techniques required to transform the short-term behaviours detected in D4.2 and D4.3 into relevant routines that may ask for specific coaching interventions.

3 Long-term Behaviour Analysis Methods

3.1 Long-term Behaviours

Long-term behaviours or routines refer to physical, social, emotional or cognitive short-term behaviours that take place over long periods of time. In contrast to the short-term behaviours that we thoroughly presented in D4.2, long-term behaviours represent more meaningful and long-lasting activities that can be used to derive information about users' lifestyles. Depending on the behaviour type and the interpretation of behaviour patterns, the period of time can range from days, weeks, months or even years.

There have been many studies using on-body and off-body sensing devices to obtain information about user's lifestyles, regarding the mobility patterns (using accelerometer and GPS data to track user's location), the sociability (using social media apps, calls, and text messages to study the interaction with other users), the cognitive behaviour (using self-reported questionnaires to monitor user's skills in different fields such as occupational and academic) and the emotional behaviour (using physiological signals, audio content and device usage patterns). However, there have been only few studies examining people's everyday behavioural patterns over time. According to (Harari, et al., 2016), smartphone data can be used to detect certain markers that characterize a person's or a group's behaviour over time and identify different types of lifestyles. For instance, monitoring long-term behaviour could reveal information about a set of patterns that can help in distinguishing a 'working lifestyle' from a 'lifestyle after retirement' (modelling between-person variations), but could also characterize subjective behavioural patterns that force individuals to have a sedentary lifestyle on specific weekdays (modelling within-person variations). Thus, long-term behaviours can reveal trends or patterns that restrains users from following a normative or healthy lifestyle.

In order to provide a clear definition for long-term behaviours, we investigated the following questions:

1. When (or how often) and why does a user deviate from a normative behaviour?
2. How does a user's behaviour manifest over time and for different units of time (e.g., hours, days, weeks)?
3. What are the main patterns or trends for long-term behaviours?
4. Are there any common patterns or trends over certain individuals or groups (e.g., diabetes patients, elderly people)?
5. Are there any individual factors that can trigger certain patterns or trends over time?
6. Are there any patterns or trends associated with psychological constructs (e.g., well-being, age, personality)?
7. How does a significant life event/intervention affect or change long-term behaviours?

The recent technological advances have allowed the researchers to integrate sensing methods in the field of psychology in order to gain knowledge about people's patterns and their lifestyles over time (Harari, et al., 2016). Even though there have been a few ongoing studies, this field is still immature. Consequently, in order to answer the aforementioned questions, we developed our own approach for modelling a long-term behaviour.

In particular, **long-term physical behaviour** can be defined based on the number of steps and/or the number of performed activities over a certain period of time (e.g., hours, days, weeks, months, years) and can be categorised as (Bassett, Toth, LaMunion, & Crouter, 2017):

1. **Sedentary:** the activity level of a user who performs <5000 steps per day (e.g., an office worker or a person who is getting no exercise)
2. **Lightly active:** the activity level of a user who performs from 5000 to 7499 steps per day (e.g., an office worker or a person who is getting little exercise)
3. **Moderately active:** the activity level of a user who performs from 7500 to 9999 steps per day (e.g., a construction worker or a person who runs one hour daily)
4. **Vigorously active:** the activity level of a user who performs >10000 steps per day (e.g., an agricultural worker or a person who runs for more than two hours daily).

Furthermore, **long-term social behaviour** can be defined based on the level of user's social interaction (short-term social behaviour) over a certain period of time (e.g., hours, days, weeks, months, years) and can be categorised as (Place, et al., 2017) (Adams, Leibbrandt, & Moon, 2011):

1. **Social withdrawal:** the average level of being socially isolated for the most part of the day (e.g., a person who is prone to avoiding people and rejecting activities who would normally enjoy)
2. **Social engagement:** the average level of being socially active for the most part of the day (e.g., a person who tends to participate in the activities of a social group that takes place in a family or in a working environment).

On the contrary and to the best of our knowledge, **long-term emotional and cognitive behaviour** cannot be clearly defined over a longer period of time. According to (Place, et al., 2017), there is not a sufficient and objective method to collect, analyse and track emotional and cognitive behaviour continuously and based on symptom-related behavioural indicators. For this reason, we decided to use our own models for detecting short-term emotional and cognitive behaviour and apply these models over longer periods of time as an attempt to identify patterns related to the user's valence/arousal (short-term emotional behaviour) and the engagement level (short-term cognitive behaviour).

A summary of the aforementioned definitions is presented in Table 1 below.

Table 1: An overview of the short-term & long-term behaviours.

Behaviours	Short-term (primitives)	Long-term (routines)
Physical	1. Steps 2. Activity Recognition: ▪ Cycling ▪ Walking ▪ Still ▪ Tilting ▪ Taking the bus (vehicle)	1. Sedentary 2. Lightly active 3. Moderately active 4. Vigorously active
Social	1. Social isolation 2. Social interaction	1. Social engagement 2. Social withdrawal
Emotional	1. Valence 2. Arousal	-
Cognitive	Level of engagement: 1. Attention 2. Distraction	-

3.2 Techniques

Different techniques are used in order to transform short-term behaviours into relevant routines and describe behavioural patterns over longer period of time. According to (Harari, et al., 2016), there have been six main techniques for describing routines over time:

1. **Unsupervised machine learning techniques:** models that learn from the test dataset which have not been previously labelled. Some of the most common techniques for unsupervised learning include clustering (e.g., k-means, mixture models, hierarchical clustering) and neural networks algorithms (Sprint, Cook, & Schmitter-Edgecombe, 2016).
2. **Time series models:** time series refer to a sequence of data points which are paired with an associated timestamp in chronological sequence. They contain valuable information about seasonal variations, health event or personal changes that might be useful when monitoring human behaviour in order to detect and analyse changes (e.g., monitoring the behaviour as a person makes progress towards a fitness goal) (Sprint, Cook, & Schmitter-Edgecombe, 2016).
3. **Psychometric methods:** it is based on psychological measurement techniques and methods related to human judgment and perception (Squires, et al., 2013).
4. **SEM-based longitudinal models:** longitudinal models are used to study the development of behaviours over time. In particular, they can be used for the analysis of change in a single/several outcomes over time (e.g., how the cognitive impairments are related to an elderly person over time or how a therapeutic intervention affects a certain behaviour over time). The Structural Equation Modelling (An, Yung, & Yang, 2015) approach is used to analyse longitudinal data based on various types of latent curve models such as the Growth Curve Modelling (GCM) and the Latent Growth Curve Modelling (LGCM) (Hox & Stoel, 2005).
5. **Change models (pre-event vs post-event):** modelling the difference of the predicted behaviour over time versus the change that actually was observed for a certain period of time (Thaler, 2005).
6. **Latent class and growth mixture modelling techniques:** latent growth modelling approaches, such as latent class growth analysis (LCGA) and growth mixture modelling (GMM) can identify homogeneous subpopulation within larger heterogeneous populations and can be used to differentiate groups and identify meaningful patterns among individuals (e.g., detect social behaviour for elderly people who live alone versus with their family or in a nursing home) (Jung & Wickrama, 2008).

Among the aforementioned techniques we will focus on time series analysis as an attempt to identify trends or patterns over time. Time series analysis can be categorised into parametric/non-parametric, linear/non-linear and univariate/multivariate. Three of the most common ways to perform time series analysis include ARIMA models, Box-Jenkins multivariate models, and Holt-Winters exponential smoothing (single, double and triple). In particular, ARIMA (Auto Regressive Integrated Moving Average) models, which are the most popular models for forecasting a time series, are based on autoregression where the data should be univariate. An overview of the existing time series forecasting methods is presented in Table 2 (Brownlee, 2018).

Table 2: An overview of the time series forecasting methods.

Time Series Method	Description	Comment
Autoregression (AR)	models the next step in the sequence as a linear function of the observations at prior time steps.	suitable for univariate time series without trend and seasonal components.
Moving Average (MA)	models the next step in the sequence as a linear function of the residual errors from a mean process at prior time steps.	suitable for univariate time series without trend and seasonal components
Autoregressive Moving Average (ARMA)	models the next step in the sequence as a linear function of the observations and residual errors at prior time steps. It combines both AR and MA models.	suitable for univariate time series without trend and seasonal components.
Autoregressive Integrated Moving Average (ARIMA)	models the next step in the sequence as a linear function of the differenced observations and residual errors at prior time steps. It combines both AR and MA models as well as the integration (I); a differencing pre-processing step of the sequence to make the sequence stationary.	suitable for univariate time series with trend and without seasonal components.
Seasonal Autoregressive Integrated Moving-Average (SARIMA)	models the next step in the sequence as a linear function of the differenced observations, errors, differenced seasonal observations, and seasonal errors at prior time steps. It combines the ARIMA model with the ability to perform the same autoregression, differencing, and moving average modelling at the seasonal level.	suitable for univariate time series with trend and/or seasonal components.
Seasonal Autoregressive Integrated Moving-Average with Exogenous Regressors (SARIMAX)	is an extension of the SARIMA model that also includes the modelling of exogenous variables.	suitable for univariate time series with trend and/or seasonal components and exogenous variables.
Vector Autoregression (VAR)	models the next step in each time series using an AR model. It is the generalization of AR to multiple parallel time series, e.g. multivariate time series.	suitable for multivariate time series without trend and seasonal components.
Vector Autoregression Moving-Average (VARMA)	models the next step in each time series using an ARMA model. It is the generalization of ARMA to multiple parallel time series, e.g. multivariate time series.	suitable for multivariate time series without trend and seasonal components
Vector Autoregression Moving-Average with Exogenous Regressors (VARMAX)	is an extension of the VARMA model that also includes the modelling of exogenous variables. It is a multivariate version of the ARMAX method.	suitable for multivariate time series without trend and seasonal components and exogenous variables.

Simple Exponential Smoothing (SES)	models the next time step as an exponentially weighted linear function of observations at prior time steps.	suitable for univariate time series without trend and seasonal components
Holt Winter's Exponential Smoothing (HWES)	models the next time step as an exponentially weighted linear function of observations at prior time steps, taking trends and seasonality into account.	suitable for univariate time series with trend and/or seasonal components
Box-Jenkins (BJ)	is a combination of the AR and Ma models and can be approximated using an ARMA model (stationary) or an ARIMA model (non-stationary)	suitable for multivariate time series with trend and seasonal components.

Furthermore, similar to the supervised classification approach that we followed in D4.2, we define here some statistical features (sum, mean, and standard deviation) to be extracted from the short-term behaviour data in order to identify potential descriptors that might reveal information about behaviour routines. These features will be used to construct meaningful time series from the short-term behaviour data, which will be later analysed by means of some of the techniques referred to above. A list of the proposed features is presented in the following tables for each behaviour.

Table 3: An overview of the proposed features to be extracted from short-term physical behaviour (steps).

Feature Name	Measurement Frequency	Description
Steps_count	hourly/daily/weekly/monthly	The number of steps performed during a certain timeframe
Steps_count_morning	daily/weekly/monthly	The number of steps performed during the morning (8am-12pm) and for a certain timeframe
Steps_count_afternoon	daily/weekly/monthly	The number of steps performed during the afternoon (12pm-5pm) and for a certain timeframe
Steps_count_evening	daily/weekly/monthly	The number of steps performed during the evening (5pm-12am) and for a certain timeframe
Steps_count_daytime	daily/weekly/monthly	The number of steps performed during the day (7am-12am)
Steps_count_work	daily/weekly/monthly	The number of steps performed during the working hours (8am-4pm) and for a certain timeframe
Steps_count_weekdays	weekly/monthly	The number of steps performed during the week days (Monday-Friday)
Steps_count_weekend	weekly/monthly	The number of steps performed during the weekend (Saturday & Sunday)
Steps_count_indoors	hourly/daily/weekly/monthly	The number of steps performed indoors and for a certain timeframe
Steps_count_outdoors	hourly/daily/weekly/monthly	The number of steps performed outdoors and for a certain timeframe

Steps_count_ratio	hourly/daily/weekly/monthly	The ratio of the performed steps for a certain timeframe compared to the previous one

Table 4: An overview of the proposed features to be extracted from short-term physical behaviour (related to all the performed activities).

Feature Name	Measurement Frequency	Description
Activity	hourly/daily/weekly/monthly	The total duration of all the performed activities during a certain timeframe
Activity_afternoon	daily/weekly/monthly	The total duration of the performed activities during the afternoon (12pm-5pm) and for a certain timeframe
Activity_evening	daily/weekly/monthly	The total duration of the performed activities during the evening (5pm-12am) and for a certain timeframe
Activity_daytime	daily/weekly/monthly	The total duration of the performed activities during the day (7am-12am)
Activity_work	daily/weekly/monthly	The total duration of the performed activities during the working hours (8am-4pm) and for a certain timeframe
Activity_weekdays	weekly/monthly	The total duration of the performed activities during the week days (Monday-Friday)
Activity_weekend	weekly/monthly	The total duration of the performed activities during the weekend (Saturday & Sunday)
Activity_indoors	hourly/daily/weekly/monthly	The total duration for the activities performed indoors and for a certain timeframe
Activity_outdoors	hourly/daily/weekly/monthly	The total duration for the activities performed outdoors and for a certain timeframe
Activity_cycling	hourly/daily/weekly/monthly	The total duration for the activity cycling and for a certain timeframe
Activity_walking	hourly/daily/weekly/monthly	The total duration for the activity walking and for a certain timeframe
Activity_still	hourly/daily/weekly/monthly	The total duration for the activity still and for a certain timeframe

Table 5: An overview of the proposed features to be extracted from short-term physical behaviour (related to the static activities being still and taking a vehicle).

Feature Name	Measurement Frequency	Description

Activity_sedentary	hourly/daily/weekly/monthly	The total duration of the static activities performed during a certain timeframe
Activity_sedentary_morning	daily/weekly/monthly	The total duration of the static activities performed during the morning (8am-12pm) and for a certain timeframe
Activity_sedentary_afternoon	daily/weekly/monthly	The total duration of the static activities performed during the afternoon (12pm-5pm) and for a certain timeframe
Activity_sedentary_evening	daily/weekly/monthly	The total duration of the static activities performed during the evening (5pm-12am) and for a certain timeframe
Activity_sedentary_daytime	daily/weekly/monthly	The total duration of the static activities performed during the day (7am-12am)
Activity_sedentary_work	daily/weekly/monthly	The total duration of the static activities performed during the working hours (8am-4pm) and for a certain timeframe
Activity_sedentary_weekdays	weekly/monthly	The total duration of the static activities performed during the week days (Monday-Friday)
Activity_sedentary_weekend	weekly/monthly	The total duration of the static activities performed during the weekend (Saturday & Sunday)
Activity_sedentary_indoors	hourly/daily/weekly/monthly	The total duration for the static activities performed indoors and for a certain timeframe
Activity_sedentary_outdoors	hourly/daily/weekly/monthly	The total duration for the static activities performed outdoors and for a certain timeframe

Table 6: An overview of the proposed features to be extracted from short-term physical behaviour (related to the dynamic activities walking and cycling).

Feature Name	Measurement Frequency	Description
Activity_light	hourly/daily/weekly/monthly	The total duration of all the performed activities that entail a light energy expenditure during a certain timeframe
Activity_moderate	hourly/daily/weekly/monthly	The total duration of all the performed activities that entail a

		moderate energy expenditure during a certain timeframe
Activity_vigorous	hourly/daily/weekly/monthly	The total duration of all the performed activities that entail a vigorous energy expenditure during a certain timeframe
Activity_light_morning	daily/weekly/monthly	The total duration of the performed light activities during the morning (8am-12pm) and for a certain timeframe
Activity_moderate_morning	daily/weekly/monthly	The total duration of the performed moderate activities during the morning (8am-12pm) and for a certain timeframe
Activity_vigorous_morning	daily/weekly/monthly	The total duration of the performed vigorous activities during the morning (8am-12pm) and for a certain timeframe
Activity_light_afternoon	daily/weekly/monthly	The total duration of the performed light activities during the afternoon (12pm-5pm) and for a certain timeframe
Activity_moderate_afternoon	daily/weekly/monthly	The total duration of the performed moderate activities during the afternoon (12pm-5pm) and for a certain timeframe
Activity_vigorous_afternoon	daily/weekly/monthly	The total duration of the performed vigorous activities during the afternoon (12pm-5pm) and for a certain timeframe
Activity_light_evening	daily/weekly/monthly	The total duration of the performed light activities during the evening (5pm-12am) and for a certain timeframe
Activity_moderate_evening	daily/weekly/monthly	The total duration of the performed moderate activities during the evening (5pm-12am) and for a certain timeframe
Activity_vigorous_evening	daily/weekly/monthly	The total duration of the performed vigorous activities during the evening (5pm-12am) and for a certain timeframe
Activity_light_daytime	daily/weekly/monthly	The total duration of the performed light activities during the day (7am-12am)

Activity_moderate_daytime	daily/weekly/monthly	The total duration of the performed moderate activities during the day (7am-12am)
Activity_vigorous_daytime	daily/weekly/monthly	The total duration of the performed vigorous activities during the day (7am-12am)
Activity_light_work	daily/weekly/monthly	The total duration of the performed light activities during the working hours (8am-4pm) and for a certain timeframe
Activity_moderate_work	daily/weekly/monthly	The total duration of the performed moderate activities during the working hours (8am-4pm) and for a certain timeframe
Activity_vigorous_work	daily/weekly/monthly	The total duration of the performed vigorous activities during the working hours (8am-4pm) and for a certain timeframe
Activity_light_weekdays	weekly/monthly	The total duration of the performed light activities during the week days (Monday-Friday)
Activity_moderate_weekdays	weekly/monthly	The total duration of the performed moderate activities during the week days (Monday-Friday)
Activity_vigorous_weekdays	weekly/monthly	The total duration of the performed vigorous activities during the week days (Monday-Friday)
Activity_light_weekend	weekly/monthly	The total duration of the performed light activities during the weekend (Saturday & Sunday)
Activity_moderate_weekend	weekly/monthly	The total duration of the performed moderate activities during the weekend (Saturday & Sunday)
Activity_vigorous_weekend	weekly/monthly	The total duration of the performed vigorous activities during the weekend (Saturday & Sunday)
Activity_light_indoors	hourly/daily/weekly/monthly	The total duration for the light activities performed indoors and for a certain timeframe
Activity_moderate_indoors	hourly/daily/weekly/monthly	The total duration for the moderate activities performed indoors and for a certain timeframe
Activity_vigorous_indoors	hourly/daily/weekly/monthly	The total duration for the vigorous activities performed indoors and for a certain timeframe

Activity_light_outdoors	hourly/daily/weekly/monthly	The total duration for the light activities performed outdoors and for a certain timeframe
Activity_moderate_outdoors	hourly/daily/weekly/monthly	The total duration for the moderate activities performed outdoors and for a certain timeframe
Activity_vigorous_outdoors	hourly/daily/weekly/monthly	The total duration for the vigorous activities performed outdoors and for a certain timeframe

Table 7: An overview of the proposed features to be extracted from short-term social behaviour (social isolation/interaction).

Feature Name	Measurement Frequency	Description
Social_interaction	hourly/daily/weekly/monthly	The total duration for being socially active during a certain timeframe
Social_isolation	hourly/daily/weekly/monthly	The total duration for being socially inactive during a certain timeframe
Social_interaction_morning	daily/weekly/monthly	The total duration for being socially active during the morning (8am-12pm) and for a certain timeframe
Social_isolation_morning	daily/weekly/monthly	The total duration for being socially inactive during the morning (8am-12pm) and for a certain timeframe
Social_interaction_afternoon	daily/weekly/monthly	The total duration for being socially active during the afternoon (12pm-5pm) and for a certain timeframe
Social_isolation_afternoon	daily/weekly/monthly	The total duration for being socially inactive during the afternoon (12pm-5pm) and for a certain timeframe
Social_interaction_evening	daily/weekly/monthly	The total duration for being socially active during the evening (5pm-12am) and for a certain timeframe
Social_isolation_evening	daily/weekly/monthly	The total duration for being socially inactive during the evening (5pm-12am) and for a certain timeframe
Social_interaction_daytime	daily/weekly/monthly	The total duration for being socially active during the day (7am-12am)
Social_isolation_daytime	daily/weekly/monthly	The total duration for being socially inactive during the day (7am-12am)
Social_interaction_work	daily/weekly/monthly	The total duration for being socially active during the working hours (8am-4pm) and for a certain timeframe

Social_isolation_work	daily/weekly/monthly	The total duration for being socially inactive during the working hours (8am-4pm) and for a certain timeframe
Social_interaction_weekdays	weekly/monthly	The total duration for being socially active during the week days (Monday-Friday)
Social_isolation_weekdays	weekly/monthly	The total duration for being socially inactive during the week days (Monday-Friday)
Social_interaction_weekend	weekly/monthly	The total duration for being socially active during the weekend (Saturday & Sunday)
Social_isolation_weekend	weekly/monthly	The total duration for being socially inactive during the weekend (Saturday & Sunday)
Social_interaction_indoors	hourly/daily/weekly/monthly	The total duration for being socially active when the location is indoors and for a certain timeframe
Social_isolation_indoors	hourly/daily/weekly/monthly	The total duration for being socially inactive when the location is indoors and for a certain timeframe
Social_interaction_outdoors	hourly/daily/weekly/monthly	The total duration for being socially active when the location is outdoors and for a certain timeframe
Social_isolation_outdoors	hourly/daily/weekly/monthly	The total duration for being socially inactive when the location is outdoors and for a certain timeframe
Social_interaction_ratio	hourly/daily/weekly/monthly	The ratio of being socially active for a certain timeframe compared to the previous one
Social_isolation_ratio	hourly/daily/weekly/monthly	The ratio of being socially inactive for a certain timeframe compared to the previous one

Table 8: An overview of the proposed features to be extracted from short-term emotional behaviour (valence/arousal).

Feature Name	Measurement Frequency	Description
Emotional_valence	hourly/daily/weekly/monthly	The valence score during a certain timeframe
Emotional_arousal	hourly/daily/weekly/monthly	The arousal score during a certain timeframe
Emotional_valence_morning	daily/weekly/monthly	The valence score during the morning (8am-12pm) and for a certain timeframe
Emotional_arousal_morning	daily/weekly/monthly	The arousal score during the morning (8am-12pm) and for a certain timeframe
Emotional_valence_afternoon	daily/weekly/monthly	The valence score during the afternoon (12pm-5pm) and for a certain timeframe
Emotional_arousal_afternoon	daily/weekly/monthly	The arousal score during the afternoon (12pm-5pm) and for a certain timeframe
Emotional_valence_evening	daily/weekly/monthly	The valence score during the evening (5pm-12am) and for a certain timeframe
Emotional_arousal_evening	daily/weekly/monthly	The arousal score during the evening (5pm-12am) and for a certain timeframe
Emotional_valence_daytime	daily/weekly/monthly	The valence score during the day (7am-12am)
Emotional_arousal_daytime	daily/weekly/monthly	The arousal score during the day (7am-12am)
Emotional_valence_work	daily/weekly/monthly	The valence score during the working hours (8am-4pm) and for a certain timeframe
Emotional_arousal_work	daily/weekly/monthly	The arousal score during the working hours (8am-4pm) and for a certain timeframe
Emotional_valence_weekdays	weekly/monthly	The valence score during the week days (Monday-Friday)
Emotional_arousal_weekdays	weekly/monthly	The arousal score during the week days (Monday-Friday)
Emotional_valence_weekend	weekly/monthly	The valence score during the weekend (Saturday & Sunday)
Emotional_arousal_weekend	weekly/monthly	The arousal score during the weekend (Saturday & Sunday)

Emotional_valence_indoors	hourly/daily/weekly/monthly	The valence score when the location is indoors and for a certain timeframe
Emotional_arousal_indoors	hourly/daily/weekly/monthly	The arousal score when the location is indoors and for a certain timeframe
Emotional_valence_outdoors	hourly/daily/weekly/monthly	The valence score when the location is outdoors and for a certain timeframe
Emotional_arousal_outdoors	hourly/daily/weekly/monthly	The arousal score when the location is outdoors and for a certain timeframe
Emotional_valence_ratio	hourly/daily/weekly/monthly	The valence score for a certain timeframe compared to the previous one
Emotional_arousal_ratio	hourly/daily/weekly/monthly	The arousal score for a certain timeframe compared to the previous one

Table 9: An overview of the proposed features to be extracted from short-term cognitive behaviour (attention/distraction).

Feature Name	Measurement Frequency	Description
Cognitive_attention	hourly/daily/weekly/monthly	The engagement level (attention score) during a certain timeframe
Cognitive_distraction	hourly/daily/weekly/monthly	The engagement level (distraction score) during a certain timeframe
Cognitive_attention_morning	daily/weekly/monthly	The engagement level (attention score) during the morning (8am-12pm) and for a certain timeframe
Cognitive_distraction_morning	daily/weekly/monthly	The engagement level (distraction score) during the morning (8am-12pm) and for a certain timeframe
Cognitive_attention_afternoon	daily/weekly/monthly	The engagement level (attention score) during the afternoon (12pm-5pm) and for a certain timeframe
Cognitive_distraction_afternoon	daily/weekly/monthly	The engagement level (distraction score) during the afternoon (12pm-5pm) and for a certain timeframe
Cognitive_attention_evening	daily/weekly/monthly	The engagement level (attention score) during the evening (5pm-12am) and for a certain timeframe

Cognitive_distraction_evening	daily/weekly/monthly	The engagement level (distraction score) during the evening (5pm-12am) and for a certain timeframe
Cognitive_attention_daytime	daily/weekly/monthly	The engagement level (attention score) during the day (7am-12am)
Cognitive_distraction_daytime	daily/weekly/monthly	The engagement level (distraction score) during the day (7am-12am)
Cognitive_attention_work	daily/weekly/monthly	The engagement level (attention score) during the working hours (8am-4pm) and for a certain timeframe
Cognitive_distraction_work	daily/weekly/monthly	The engagement level (distraction score) during the working hours (8am-4pm) and for a certain timeframe
Cognitive_attention_weekdays	weekly/monthly	The engagement level (attention score) during the week days (Monday-Friday)
Cognitive_distraction_weekdays	weekly/monthly	The engagement level (distraction score) during the week days (Monday-Friday)
Cognitive_attention_weekend	weekly/monthly	The engagement level (attention score) during the weekend (Saturday & Sunday)
Cognitive_distraction_weekend	weekly/monthly	The engagement level (distraction score) during the weekend (Saturday & Sunday)
Cognitive_attention_indoors	hourly/daily/weekly/monthly	The engagement level (attention score) when the location is indoors and for a certain timeframe
Cognitive_distraction_indoors	hourly/daily/weekly/monthly	The engagement level (distraction score) when the location is indoors and for a certain timeframe
Cognitive_attention_outdoors	hourly/daily/weekly/monthly	The engagement level (attention score) when the location is outdoors and for a certain timeframe
Cognitive_distraction_outdoors	hourly/daily/weekly/monthly	The engagement level (distraction score) when the location is outdoors and for a certain timeframe
Cognitive_attention_ratio	hourly/daily/weekly/monthly	The engagement level (attention score) for a certain timeframe compared to the previous one

Cognitive_distraction_ratio	hourly/daily/weekly/monthly	The engagement level (distraction score) for a certain timeframe compared to the previous one

4 Evaluation

Overall, a time series consists of systematic and non-systematic components. Systematic components have consistency or recurrence and can be described and modelled. These components include the level (the average value in the series), the trend (the increasing/decreasing value in the series) and the seasonality (the repeating short-term cycle in the series). On the other hand, non-systematic components, called the noise/irregularity (random variation) in the time series, cannot be directly modelled. Thus, a time series can be represented as an aggregate or combination of these four components. Furthermore, each one of these components can be depicted through the time series decomposition function. The `statsmodels` library in python provides an implementation of the decomposition method in a function called `seasonal_decompose()` (Perktold, Seabold, & Taylor, 2019). The decomposition function is primarily used for time series analysis as a tool to decompose model's complexity and capture each of these components in the given model.

As a preliminary evaluation of our approach for the modelling of long-term behaviours, we collected accelerometer data for one month and we calculated the steps for every hour (short-term physical behaviour). The data was acquired for one participant (N=1) and for a period of 4 weeks (from 7 January 2019 to 3 February 2019) through a smartphone device. The participant was allowed to use the smartphone as usual and no further instructions were given.

Based on the 'Steps_count' feature, we computed the sum of the steps time series for every day and we tried to identify any trend over the period of 4 weeks. The overview of the performed steps is depicted in Figure 1. Initially, it can be seen that the highest peaks are performed during the first two weekends, and especially on Saturdays. Furthermore, we can assume that the user is more physically active on Saturdays compared to the rest of the week days.

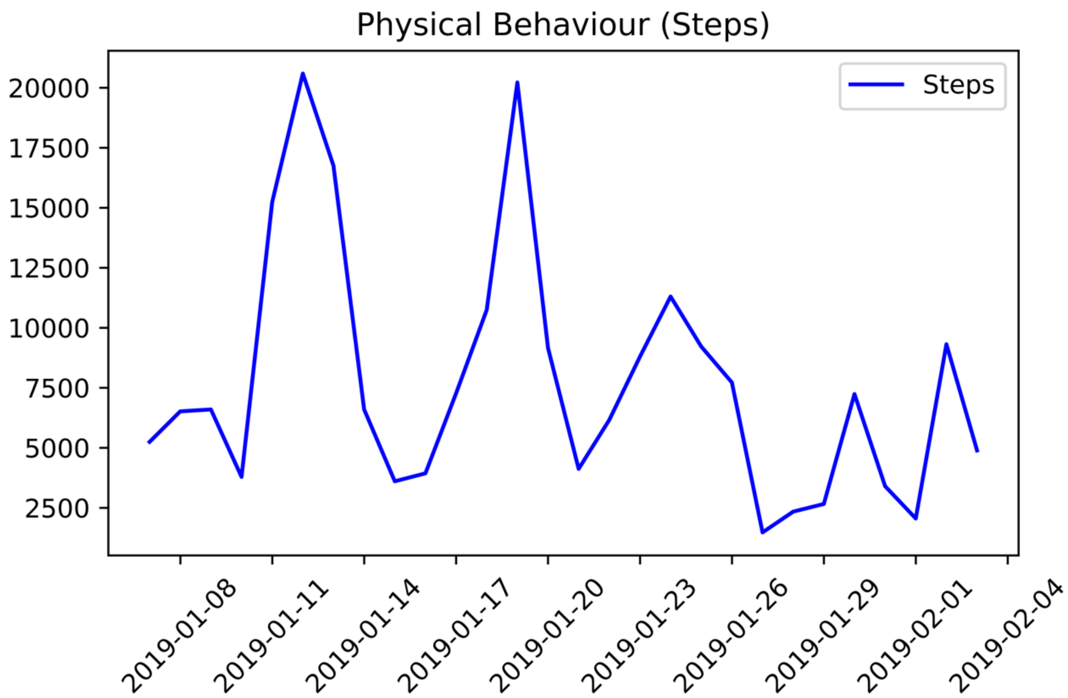

Figure 1: Long-term physical behaviour (steps) for 1 month.

Then, we decomposed the time series components using moving averages, and assuming that the series is an additive model; an additive model suggests that the four abovementioned components are added together. Thus, we can see in Figure 2 a down-trend, as the number of steps is decreasing after the second week.

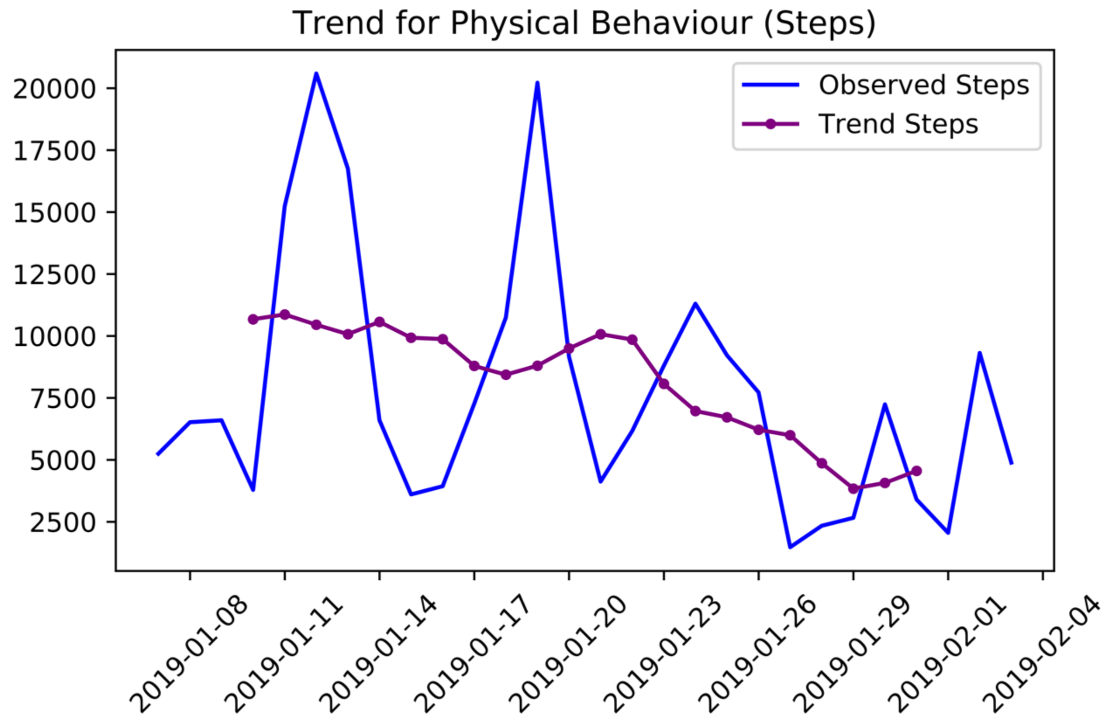

Figure 2: Long-term physical behaviour (steps) trend for 1 month.

Finally, we calculated the weighted average in order to see the trends inside the time series. For this reason, we computed the mean and standard deviation of the series for each week (the rolling size is 7) and we converted the time series from non-stationary to stationary. It is clear in Figure 3 that there is an abnormal behaviour after the '2019-01-21'. After careful consideration of all the relevant facts, we came up to the conclusion that the time series are affected by the 'seasonality', since the data collection started some days after the Christmas holidays. In particular, the irregularity of the data on Monday 21 (where a significant down-trend is noticed) is caused due to the 'Blue Monday'. It is worth mentioning that 'Blue Monday' (the 3rd Monday in January) has been identified as the most depressing day of the year, where people are prone to getting socially isolated and being physically inactive. In T4.4, we will try to gain some further insights of the long-term behaviour, by also analysing short-term social, emotional and cognitive behaviours.

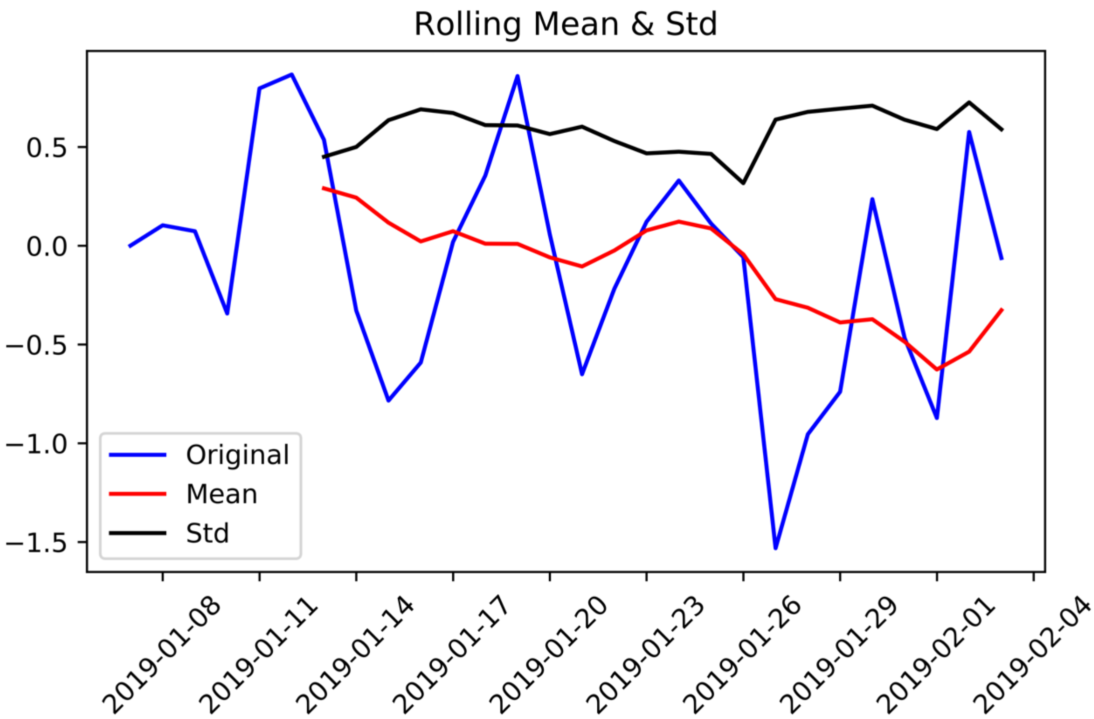

Figure 3: Long-term physical behaviour (steps) for 1 month with stationary time series.

5 Discussion

5.1 Main Findings

In the previous sections we described the methods for the inference of long-term behaviours based on short-term behaviours. Due to the limited number of related works in describing behavioural patterns over time, we came up with our own definition and approach for the long-term behaviours.

For the long-term physical behaviour, we scaled the behavioural patterns to sedentary, lightly active, moderately active and vigorously active. Statistical features are to be extracted from the short-term physical behaviour, based on steps (see Table 3) and the detected activities (see Table 4). These features will be used in D4.5 to classify the aforementioned routines over a longer period of time (hours, days, weeks, and months). Similarly, regarding the long-term social behaviour we scaled the social routines to social engagement and social withdrawal, considering the detected short-term social behaviours. Regarding the social behaviour, emotional and cognitive behaviour more data need to be collected in order to be able to identify routines or trends for longer periods of time.

5.2 Open Issues

Despite an actual evaluation of the proposed methods is not aimed at this deliverable, we considered proper to prove our concept via an exemplary study. We decided to focus on physical behaviour and specifically on steps due to its relevance for the considered use cases. It must be noted though that no clear conclusions can be drawn from this study due to the limited sample set (N=1) and duration of the study (one month).

The methods presented in section 3 & 4 are subject to improvement with the course of the planned future evaluations (T4.4). Currently, we are conducting a data collection in order to evaluate the long-term physical behaviour, including the routines for social behaviour as well. In this work-in-progress, we are collecting data for a longer period of time and for more participants in order to have a more clear understanding of the pros/cons of our approach. The new findings will be reported in upcoming deliverables of WP4.

6 Bibliography

Adams, K. B., Leibbrandt, S., & Moon, H. (2011). A critical review of the literature on social and leisure activity and wellbeing in later life. *Ageing & Society, 31(4)*, 683-712.

An, X., Yung, Y., & Yang, Q. (2015). The SEM Approach to Longitudinal Data Analysis Using the CALIS Procedure. *SAS Global Forum 2015*.

Bassett, D. J., Toth, L., LaMunion, S., & Crouter, S. (2017). Step Counting: A Review of Measurement Considerations and Health-Related Applications. . *Sports Med* .

Brownlee, J. (2018, August 6). *11 Classical Time Series Forecasting Methods in Python (Cheat Sheet)*. Retrieved from Machine Learning Mastery: https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/

Harari, G., Lane, N., Wang, R., Crosier, B., Campbell, A., & Gosling, S. (2016). Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges. *Perspect. Psychol. Sci.*

Hox, J., & Stoel, R. (2005). Multilevel and SEM Approaches to Growth Curve Modeling. *Encyclopedia of Statistics in Behavioral Science, Volume 3, John Wiley & Sons, Ltd,*, 1296–1305.

Jung, T., & Wickrama, K. (2008). An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass 2/1*, 302–317.

Perktold, J., Seabold, S., & Taylor, J. (2019, February 15). *statsmodels.tsa.seasonal.seasonal_decompose*. Retrieved from statsmodels: http://www.statsmodels.org/dev/generated/statsmodels.tsa.seasonal.seasonal_decompose.html

Place, S., Blanch-Hartigan, D., Rubin, C., Gorrostieta, C., Mead, C., Kane, J., . . . Azarbayejani, A. (2017). Behavioral Indicators on a Mobile Sensing Platform Predict Clinically Validated Psychiatric Symptoms of Mood and Anxiety Disorders. *J Med Internet Res.* .

Sprint, G., Cook, D., & Schmitter-Edgecombe, M. (2016). Unsupervised detection and analysis of changes in everyday physical activity data. *Journal of Biomedical Informatics, Volume 63*, 54-65.

Squires, J., Hayduk, L., Hutchinson, A., Cranley, L., Gierl, M., Cummings, G., . . . Estabrooks, C. (2013). A protocol for advanced psychometric assessment of surveys. . *Nursing Research and Practice, vol. 2013, Article ID 156782*.

Thaler, R. H. (2005). *Advances in Behavioral Finance, Volume II*. Oxford: Princeton University Press.