D4.6: Methods for detecting behaviour changes from short-term behaviour information

Dissemination level: Public
Document type: Report
Version: 1.0.0
Date: November 29, 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 detecting behaviour changes from short-term behaviour information
Due date of deliverable	November 30, 2019
Work package	WP4
Author(s)	Oresti Banos (CMC), Kostas Konsolakis (CMC)
Reviewer(s)	Alison Pease (UDun), Harm op den Akker (RRD), Jorien van Loon (CMC)
Approved by	Coordinator
Dissemination level	Public
Document type	Report
Total number of pages	37

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 (D4.6) represents the contribution of Work Package 4 (WP4) to the automatic detection of user's behaviour changes. Such changes are investigated, as part of task T4.3, in order to discover any relevant alterations in the user lifestyle. The behaviour changes are automatically detected from the analysis of the short-term behaviour time series generated in T4.2. The detected changes, and their relevance, are then communicated to the council to trigger potential interventions. Thus, this deliverable aims to research and develop the methods for automatically detecting and assessing behaviours changes.

Table of Contents

List of figures

List of tables

Symbols, abbreviations and acronyms

BCSS	Behaviour Change Support System
CaaS	Coach-as-a-Sensor
CMC	Centre for Monitoring and Coaching
COUCH	Council of Coaches
CPD	Change Point Detection
D	Deliverable
EC	European Commission
ESM	Experience Sampling Methods
HBAF	Holistic Behaviour Analysis Framework
HCI	Human-Computer Interaction
IBC	Interventional Behaviour Changes
M	Month
MS	Milestone
OBC	Observational Behaviour Changes
PELT	Pruned Exact Linear Time
RRD	Roessingh Research and Development
SKB	Shared Knowledge Base
SU	Sorbonne University
T	Task
TP	True Positives
WP	Work Package
UPV	Universitat Politècnica de València
UT	University of Twente

1 Introduction

After researching the inference of short-term and long-term behaviours from sensor data (T4.1 and T4.2 respectively), T4.3 comes into place to explore potential variations in the measured behaviours. Hereafter termed as "behaviour changes", these variations represent potential deviations from the user's normal behaviour patterns or routines, which may in some cases demand a prompt or preventive intervention from a specific coach or the whole council. To that end, the time series generated from the sequence of everyday short-term behaviours estimated from the user are processed and analysed to detect and quantify relevant fluctuations or tipping points. Hence, the analysis consists in detecting changes between time periods (change detection), determining the significance of the detected changes (change assessment) and analysing the nature of the change (change explanation). The detection and assessment of behaviour changes is approached via the automatic analysis of the time series generated from users' daily short-term behaviours. The explanation of the changes is rather aimed to be attained via the user-council conversations, hence not specifically addressed in this deliverable. Finally, the detected changes can be made available to the council to trigger specific coaching actions as defined in T3.2 and executed by the Dialogue and Argumentation Framework in WP5. All in all, this deliverable aims to describe the methodology developed for detecting and quantifying user behaviour changes.

The deliverable is structured as follows. Section 2 presents the main objectives of this deliverable. Section 3 provides an overview on the user's behavioural information to be considered for detection analysis. Hence, this section recaps on the physical and social behaviour information generated from sensors and also introduces the couch-as-a-sensor concept and its utility for measuring emotional and cognitive behaviours. Section 4 describes the methods used for detecting behaviour changes, including the related terminology and the techniques for behaviour change analysis. Section 5 presents a preliminary evaluation of these methods for detecting behaviour changes. Finally, Section 6 discusses the main outcomes of this deliverable.

2 Objectives

The main objective of this deliverable (D4.6) is to describe the methods developed for the detection of behaviour changes in the user's behaviour patterns. 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, and also to a certain extent the long-term behaviours, detected in D4.4 and D4.5, into relevant estimators of behaviour changes demanding specific coaching interventions.

3 Short-term Behaviour Information

Before describing the methodology for estimating and assessing behaviour changes, which is the primary purpose of D4.6, this section provides an overview of the information generated by the Holistic Behaviour Analysis Framework (HBAF) for the user's short-term physical, social, emotional and cognitive behaviours. As it was defined in D4.2, short-term behaviours represent behaviours that last less than a certain period of time (e.g. hours/days). Hence, the behaviour data points measured every day define the time series that will serve as input to the behaviour change detection methods.

Short-term physical behaviour consists of the user's number of steps but also the duration of each activity performed by the user. Sedentary activities consist of sitting, standing, tilting (e.g. from sitting to standing) and commuting (e.g. driving, taking the bus), while vigorous activities consist of walking, running and cycling. **Short-term social behaviour** consists of the user's duration of social activation or isolation. More information about the methodology developed to transform raw sensor data into short-term physical and social behaviours can be found in deliverable D4.2.

Based on the findings of D4.1, there are just a few ongoing studies aiming to measure emotional and cognitive behaviour from raw smartphone and smartwatch data. According to our previous work (Konsolakis, Hermens, Villalonga, Vollenbroek-Hutten, & Banos, 2018) and to the best of our knowledge, there is not an accurate approach for predicting user's emotional and cognitive behaviour from sensor data, especially when it comes to generating a daily overview of the user emotional and cognitive state. As a result, we decided to adopt a more effective way of acquiring this type of information through experience sampling method (ESM), which is here evolved by means of the Coach-as-a-Sensor concept.

The **Coach-as-a-Sensor** (CaaS) represents any type of information that can be actively acquired during the Council of Coaches sessions and not passively through sensor data. During the regular Council of Coaches sessions, the user is asked some questions in order to monitor user's emotional and cognitive behaviour. Such questions are interspersed with the normal flow of the conversation, and in some cases, may be regarded as small talk or conversational connectors, thus reducing the perceived burden to the user compared to traditional ESM's. Thus, for example, the Peer Coach (Carlos, see D3.4) will ask questions related to the user's mood (e.g., "*How are you feeling today? Could you describe your overall mood during the morning hours?*" as depicted in Figure 1). Similarly, the Cognitive Coach (Helen) asks questions related to the engagement of the user with cognitive tasks (e.g., "*Have you read a book today?*" as depicted in Figure 2). Both of these coaches aim to estimate the total duration of being sad (including answers for being very negative or negative), happy (including answers for being positive or very positive) or involved with cognitive tasks. An overview of the questions asked through the Coach-as-a-Sensor is presented in Table 1.

Table 1: Questionnaires for monitoring user's emotional and cognitive behaviour.

	Question	Answer
Emotional Behaviour	*How are you feeling today? Please describe your overall mood (from negative to positive) during morning hours (8am-12pm)*	▪ Very negative ▪ Negative ▪ Neutral ▪ Positive ▪ Very Positive
	Please describe your overall mood (from negative to positive) during afternoon hours (12pm-5pm)	▪ Very negative ▪ Negative ▪ Neutral ▪ Positive ▪ Very Positive
	Please describe your overall mood (from negative to positive) during evening hours (5pm-12am)	▪ Very negative ▪ Negative ▪ Neutral ▪ Positive

		▪ Very Positive
Cognitive Behaviour	*Have you participated today in any of the following tasks, such as reading a book/newspaper, acquiring a new skill, watching an educational TV-show, or playing a board game/sudoku? If so, please mention the total duration in minutes (for every given answer.*	User's input for the total duration in minutes

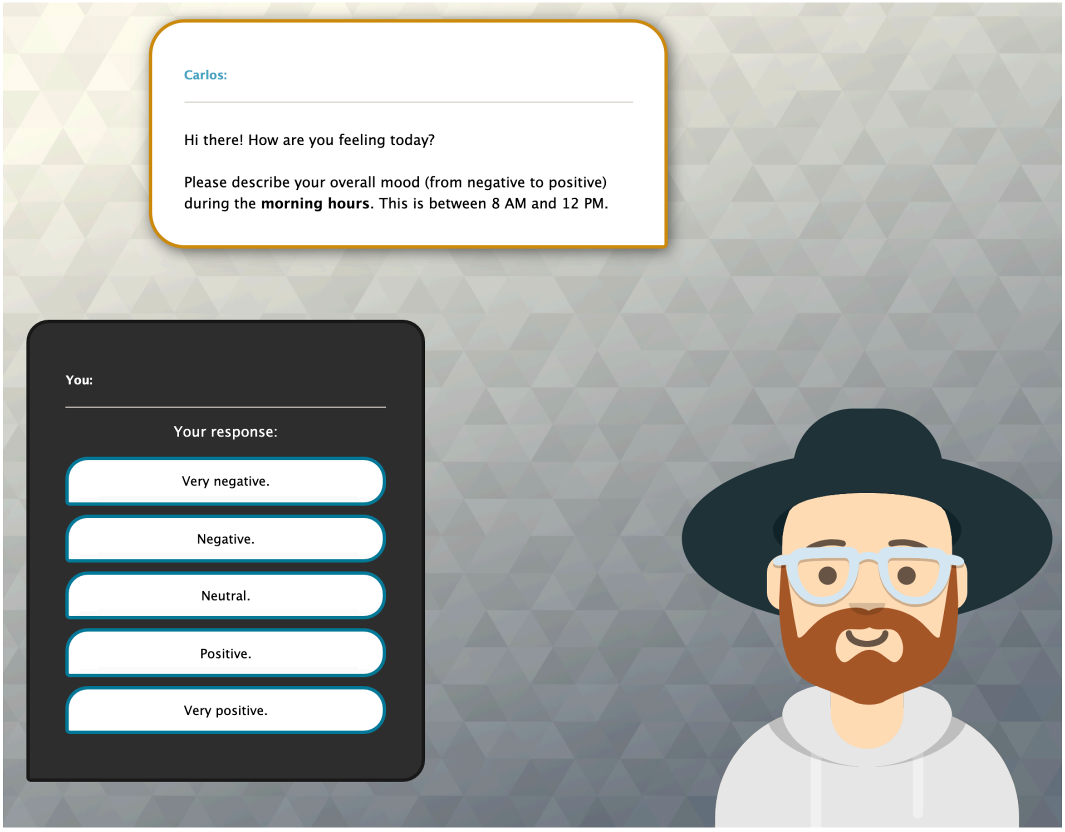

Figure 1: Coach-as-a-Sensor: monitoring emotional wellbeing (example using the WOOL Dialogue Framework editor that is used to author and test dialogues).

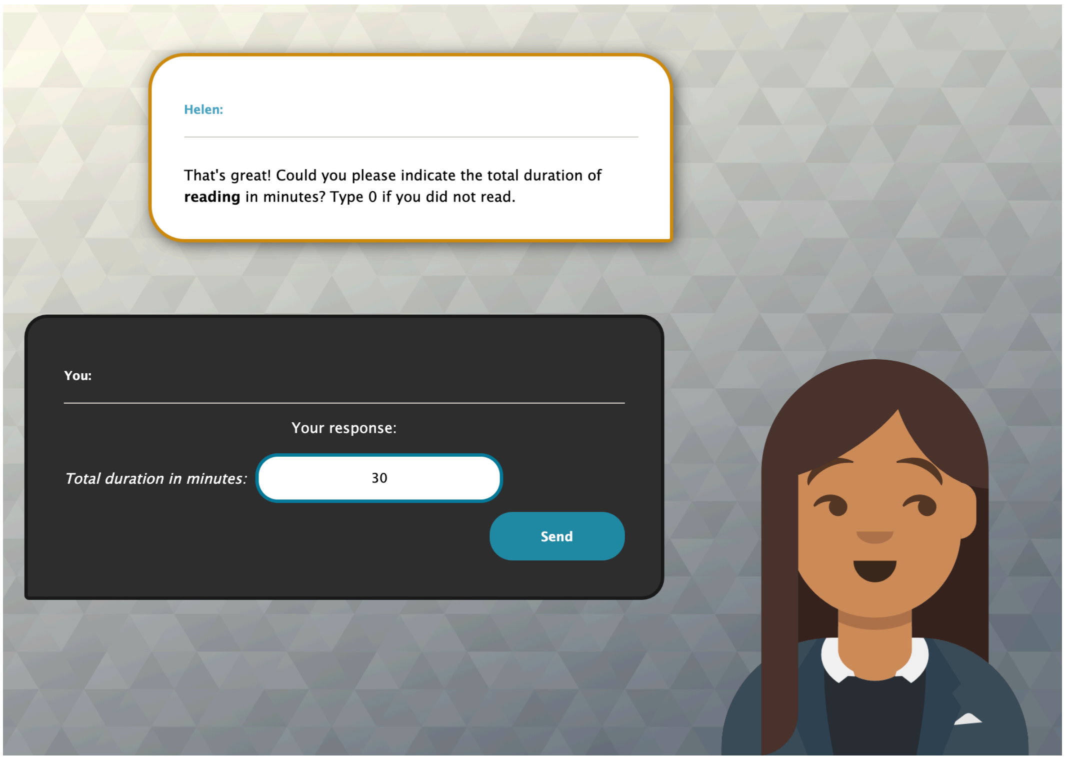

Figure 2: Coach-as-a-Sensor: monitoring cognitive behaviour (example using the WOOL Dialogue Framework editor that is used to author and test dialogues).

It is worth mentioning that the emotional and cognitive short-term behaviour data acquired through the CaaS is communicated to the HBAF via the Shared Knowledge Base (SKB). The mechanism is described in more detail in D4.7 (see Section 3.2).

An overview of the short-term behaviour information available for modelling the behaviour changes is depicted in Figure 3. Note that each column represents a time-series.

short_term_behaviour - DataFrame

Index	Day	Week	Steps	Sedentary_Duration	Vigorous_Duration	Social_Interaction	Social_Isolation	Happiness_Duration	Sadness_Duration	Cognitive_Duration
2019-09-30 00:00:00	Monday	4	8002	1265	175	475	965	105	150	0
2019-10-01 00:00:00	Tuesday	4	5670	1330	110	435	1005	135	60	0
2019-10-02 00:00:00	Wednesday	4	6002	1320	120	465	975	120	110	45
2019-10-03 00:00:00	Thursday	4	7050	1290	150	485	955	105	85	0
2019-10-04 00:00:00	Friday	4	8400	1260	180	540	900	180	45	0
2019-10-05 00:00:00	Saturday	4	4150	1365	75	375	1065	265	60	45
2019-10-06 00:00:00	Sunday	4	3010	1395	45	345	1095	240	120	60
2019-10-07 00:00:00	Monday	5	4003	1380	60	270	1170	105	110	120
2019-10-08 00:00:00	Tuesday	5	3270	1395	45	245	1195	120	60	0
2019-10-09 00:00:00	Wednesday	5	3500	1385	55	295	1145	135	90	60
2019-10-10 00:00:00	Thursday	5	3900	1385	55	240	1200	105	85	0
2019-10-11 00:00:00	Friday	5	4900	1340	100	295	1145	165	75	0
2019-10-12 00:00:00	Saturday	5	3000	1390	50	355	1085	240	60	0
2019-10-13 00:00:00	Sunday	5	1500	1420	20	315	1125	210	120	75
2019-10-14 00:00:00	Monday	6	7003	1290	150	310	1130	150	75	15
2019-10-15 00:00:00	Tuesday	6	5550	1330	110	295	1145	180	45	0
2019-10-16 00:00:00	Wednesday	6	6500	1300	140	305	1135	165	75	60
2019-10-17 00:00:00	Thursday	6	7055	1285	155	300	1140	180	60	45
2019-10-18 00:00:00	Friday	6	6900	1295	145	325	1115	210	60	30
2019-10-19 00:00:00	Saturday	6	7100	1285	155	355	1085	240	45	60
2019-10-20 00:00:00	Sunday	6	6150	1310	130	345	1095	225	90	90

Figure 3: Example of short-term behaviour dataset.

4 Methods for detecting behaviour changes

4.1 Behaviour Changes

Human behaviours are related to habits, actions and responses to a given situation or stimulus. Behaviour changes can be defined as deviations from the normal realisation of activities, interactions or attitudes, which show in a new or different way compared to the regular lifestyle of a person. These changes might be either the outcome of a specific intervention aiming to change behaviour or the result of an unforeseen event that occurs naturally. In the former case, behaviour changes can occur during the course of a specific treatment, where patients and physicians set out new goals representing new behaviours or routines to be followed by the patient. For example, patients with diabetes might be asked to change their activity habits and be more active, guided by specific instructions such as walking a minimum number of steps per day in order to lose weight. On the other hand, behaviour change can take place naturally without any interventions taking place. An example of this might be a patient with diabetes who shows a drop in physical activity or sleep hours due to seasonal irregularities, such as national holidays, vacations, etc. In the context of this work, we are particularly interested in the observation and analysis of daily and weekly behaviour changes, irrespective of whether they are driven by an intervention or not.

4.2 Behaviour Change Detection Techniques

While there is whole realm of research dealing with the definition and development of mechanisms to induce behaviour change, there is a clear lack of techniques addressing the automatic detection of such changes. Hence, this work aims at help filling this gap with new approaches to detect and quantify behaviour changes. The idea revolves around the analysis of the long-term behaviours (T4.2), which are, in practice, defined via time series of short-term behaviours (T4.1). The techniques proposed here intend to detect transitions or tipping points in these time series (change detection), normally in the range of days or weeks, and also quantify the relevance of the detected changes (change assessment). The techniques proposed here are thus fairly inspired by the methodology used in the time series analysis domain. A detailed description of these techniques is provided in the following paragraphs.

The detection of changes in a time series is normally referred to as Change Point Detection (CPD). Normally, statistical characteristics of the time series (e.g. mean, standard deviation) are compared for different segments of the series to determine differences. In our case, this means that the original time series generated from the computed short-term behaviours needs to be divided into a series of non-overlapping segments. In Figure 4, there is a CPD example of a user's short-term physical behaviour. Each segment corresponds to a certain physical behaviour pattern with its own homogeneous characteristics and features. Hence, the challenge lies in finding an appropriate mechanism to define the temporal boundaries of those segments, which in turn is what the CPD methods are intended for.

CPD methods detect abrupt shifts in time series trends (i.e. shifts in a time series' instantaneous velocity), that can be easily identified via the human eye, but are hard to pinpoint using traditional statistical approaches. CPD can be used for detecting anomalous sequences or states in a time series, detecting the average velocity of unique states in a time series and detecting a sudden change in a time series state in real-time.

There are two different categories of CPD: offline and online methods. On the one hand, online methods aim to detect changes on live-streaming time series as soon as they occur, usually for the purpose of constant monitoring or immediate anomaly detection. On the other hand, offline methods (also known as retrospective or posteriori) detect changes retrospectively (no use of live-streaming data) and require the complete time series in order to apply statistical analysis. Because offline approaches analyse the whole time series and perform segmentation after the time series has been collected, they are considered more accurate. While online methods can be of interest for noticing critical events (normally related to clinical exacerbations), in our case, the offline methods are favoured since the intention is to have a comprehensive picture of the user behaviour over time as to avoid possible misjudgements. In this section, we propose a number of algorithms for the detection of multiple change points in multivariate time series based on literature.

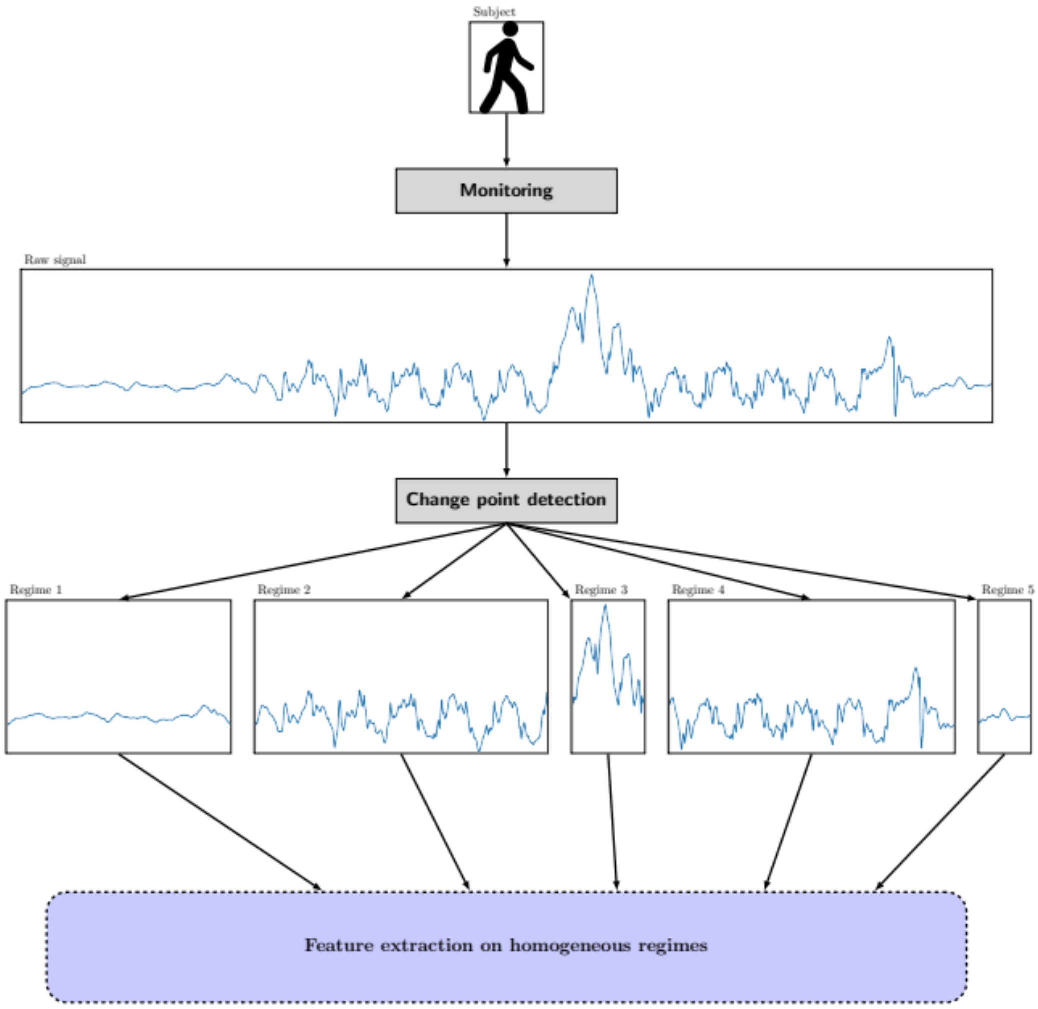

Figure 4: Flowchart of CPD for physical behaviour (Truong, Oudre, & Vayatis, 2020).

There are a number of different CPD algorithms with varying strengths and weaknesses. Truong et al. (Truong, Oudre, & Vayatis, 2020) categorize these algorithms based on three elements: cost function, search method, and constraint. The cost function measures the homogeneity of the data points, where it gets lower when data are homogenous (no change points) and larger when data are heterogeneous (it contains one or several change points). The search method yields a solution, either optimally or approximately, to the discrete optimization problem of defining the change points. The constraint element on the number of change points is added in the form of a complexity penalty when the number of changes is unknown. A larger penalty can be used to detect only the most significant changes (or even none), while a smaller penalty detects many change points.

Search method algorithms strike a balance between accuracy and computational complexity, and thus, have been investigated in literature the most. The typology of the search methods is depicted in Figure 5. They are categorized into optimal detection of change points, including the Dynamic Programming and the Pruned Exact Linear Time (PELT) algorithms, and into the approximate detection of change points, including the Binary Segmentation, Bottom-up Segmentation and the Window-sliding change point detection algorithms (Truong, Oudre, & Vayatis, 2020).

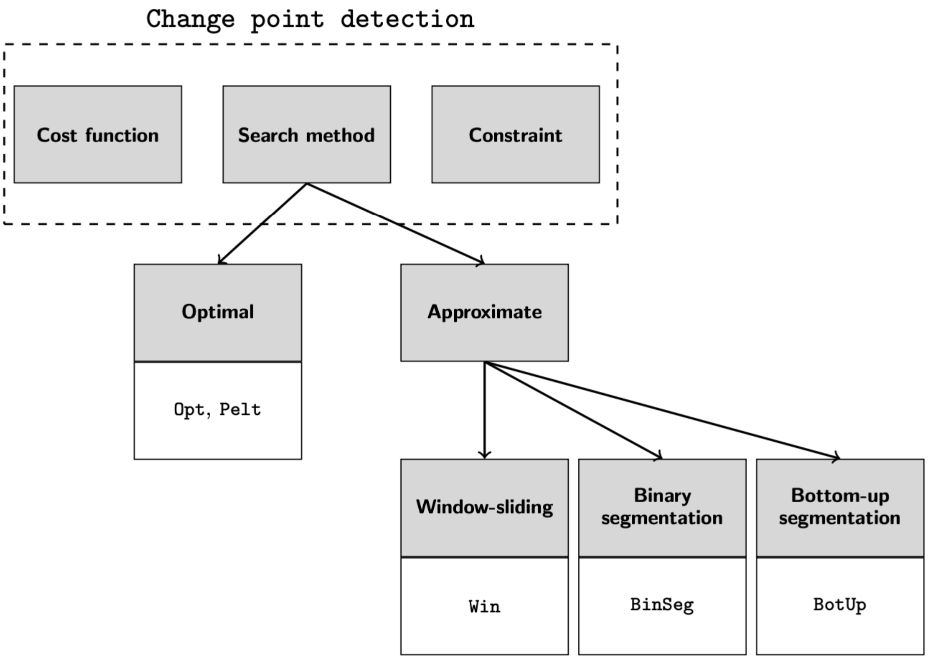

Figure 5: Typology of the CPD search methods (Truong, Oudre, & Vayatis, 2020).

Dynamic programming CPD search method: This is an exact segmentation method, which has a considerable computational cost of $O(Kn^2)$, where K is the max number of change points and n is the number of data points. This method is based on dynamic programming in order to find the optimal change points. Specifically, it computes the best partition for which the sum of errors is minimum for a given segmentation.

Pruned Exact Linear Time (PELT) CPD search method: The PELT is an exact segmentation method and generally produces quick and consistent results. It detects change points through the minimization of costs. For a given model and penalty level, it computes the optimal segmentation with the minimum approximation errors. The algorithm has a computational cost of O(n), where n is the number of data points.

Window-sliding CPD search method: This method is quite fast and simple, using two windows which slide along the data stream. The window-based search method compares the statistical properties of the time series with a discrepancy measure. When the two windows are highly dissimilar, the discrepancy reaches large values resulting into a peak which is indicative of a change point. Once the discrepancy curve has been generated, the algorithm locates optimal change point indexes in the sequence.

Binary segmentation CPD search method: This method is used to perform fast segmentation and it is possibly the most used in the literature. Binary segmentation is an approximate method with an efficient computational cost of O(n log n), where n is the number of data points. The algorithm works by iteratively applying a single change point method to the entire sequence to determine if a split exists. If a split is detected, then the sequence splits into two sub-sequences. Then, the same process is repeated on the two resulting sub-sequences.

Bottom-up segmentation CPD search method: Similarly to the binary segmentation, this method is used to perform fast segmentation with computational cost of O(n log n), where n is the number of data points. In contrast to the binary segmentation, which is a greedy procedure, bottom-up segmentation starts with a maximum number of change points and then, successively, deletes the less significant ones. At first, the time series is divided into many sub-time series which results into several change points. Then, the divided segments are merged based on their level of homogeneity, concluding with the most significant change points.

5 Evaluation

A preliminary evaluation of the aforementioned CPD methods is shown in this section in order to compare their usefulness to detect behaviour change. To that end, appropriate performance metrics need to be computed and assessed empirically on a given dataset. Hence, we evaluate the performance of the CPD methods using two different datasets, which represent short-term behaviour data for a controlled (Experiment 1) and uncontrolled scenario (Experiment 2). The data for the controlled scenario is created synthetically as to simulate a set of well-known behaviour changes for a period of two months. This allows us to characterise each CPD method with a clear prior knowledge on when and for how long the changes take place. The data of the uncontrolled scenario, which also represent short-term behavioural data for two months, is rather considered to obtain a very first indication of the utility of the selected CPD method when used in the wild. The data of the uncontrolled scenario was collected in a realistic setting and with sensor devices similar to the ones used in this project. This evaluation does not replace the formal evaluation originally planned for T4.4. Finally, the implementation of the evaluated CPD methods is available in D4.7.

Some of the most common metrics used in literature for CPD are presented next in order to evaluate the segmentation performance (Truong, Oudre, & Vayatis, 2020). For the following metrics we assume that the set of true change points is denoted by $T^* = \{t^*_1, \ldots, t^*_K\}$, while the set of estimated change points is denoted by $\hat{T} = \{\hat{t}_1, \ldots, \hat{t}_K\}$:

Precision and Recall: measure when a true CPD occurs. Given a margin M, true positives TP are defined as the true change points for which there is an estimated one at less than M samples. Precision is the proportion of predicted change points that are true change points, while recall is the proportion of true change points that are well predicted.

$$Precision = |TP| / |\{\hat{t}_l\}_l| \quad \text{and } Recall = |TP| / |\{t_k\}_k|$$

$$TP = \{ t_k \mid \exists\ \hat{t}_l \text{ s.t. } |\hat{t}_l - t_k| < M\}$$

Hausdorff metric: measures the robustness of the CPD methods. To that end, this metric estimates the worst prediction error. It is defined as the greatest temporal distance between an actual change point and its estimation:

$$Hausdorff\,(T^*, \hat{T}) = \max\{ \max_k \min_l |\, t_k - \hat{t}_l \,|\, ,\ \max_k \min_l |\, \hat{t}_k - t_l \,|\}$$

Rand index: represents the average similarity between the estimated breakpoint set $\hat{T}$ and the true T^*, and measures the number of agreements between two segmentations. For a time series $\{y_t\}_t$ and a segmentation S, and let A be (the associated adjacency matrix), where:

$A_{ij} = 1$ *if both samples* y_i *and* y_j *are in the same segment according to S, and*

$A_{ij} = 0$ *for otherwise*

Let $\hat{S}$ be the estimated segmentation, $\hat{A}$ the associated adjacency matrix and T the number of samples. Then the Rand index has a value between 0 and 1, where 0 indicates that the two segmentations do agree on any pair of points while 1 indicates that the two segmentations are exactly the same. The Rand index is equal to:

$$Rand = \frac{\sum_{i<j}(A_{ij} = \hat{A}_{ij})}{T(T-1)/2}$$

A schematic representation of the above metrics is presented in Figure 6.

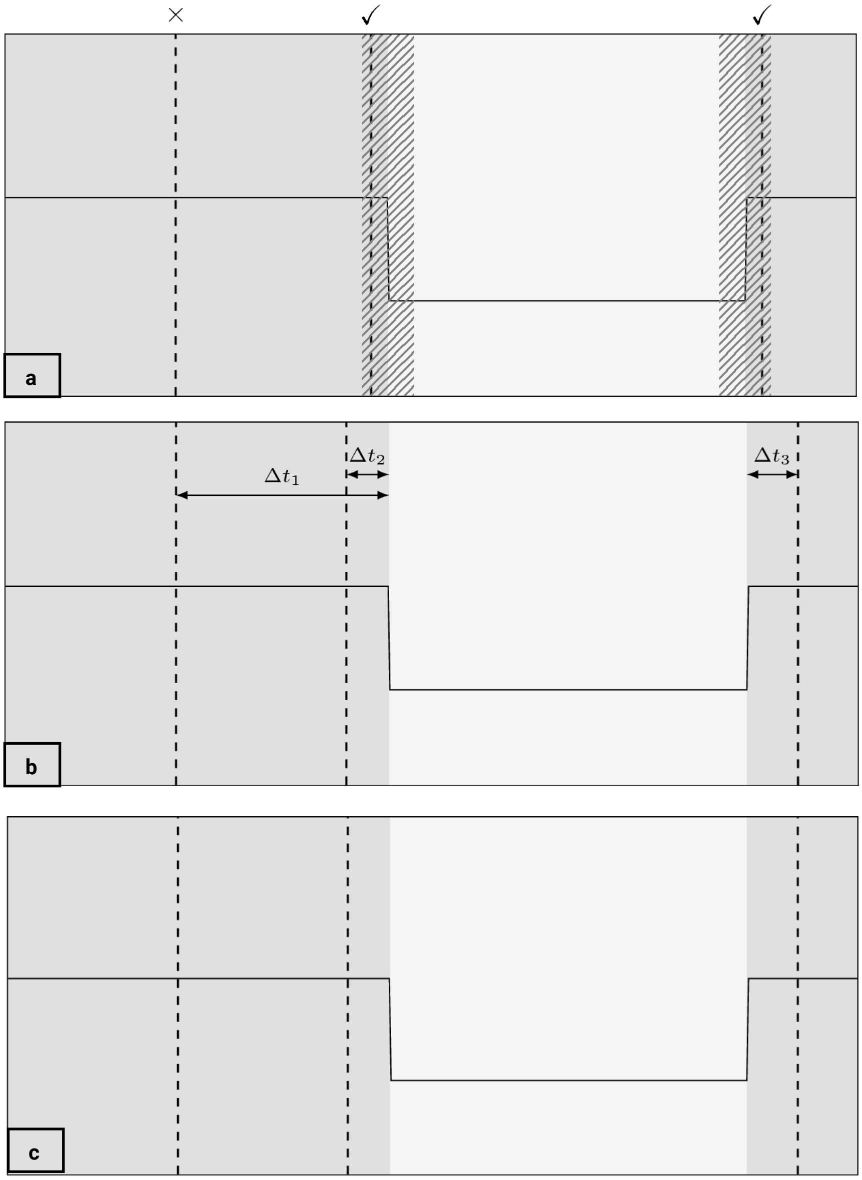

Figure 6: Schematic example where alternating grey areas mark the true segmentation T* and dashed lines mark the estimated segmentation $\widehat{T}$: a) dashed areas mark the allowed margin of error around true change points, where precision is 2/3 and recall is 2/2; b) Hausdorff is equal to $\Delta t_1 = \max(\Delta t_1, \Delta t_2, \Delta t_3)$; c) Rand index is equal to 1 minus the grey area (Truong, Oudre, & Vayatis, 2020).

5.1 Experiment 1: Holistic Controlled Dataset

This dataset represents data generated synthetically for a predefined controlled scenario. Specifically, we follow a holistic approach where short-term physical, social, emotional, and cognitive behavioural data have been simulated for a period of eight weeks.

5.1.1 Experimental Setup

In order to simulate the data, we developed the following scenario. Let us consider Michael, a 66 years old man, who works in a financial department, who is about to retire on the 7th of October 2019. As the retirement represents a big challenge in his life, Michael decides to use the COUCH system. His main goal is to have a smooth adaptation to 'life after retirement', by receiving advice through the 'Council of Coaches' system.

During the four weeks that precede retirement (9/9/2019 – 6/10/2019), Michael has been quite busy with his office duties trying to set up the tasks to his successors. On a normal weekday he works around 8 hours. After work, he goes home and spends the rest of the day with his family. An average number of steps for him, during a working day, is 7000. On the other hand, during weekends he prefers to stay at home and relax, and thus, he tends to have a rather sedentary lifestyle on Saturdays and especially on Sundays. Likewise, Michael is more socially active during working days, where he also interacts with his colleagues. His emotional behaviour is mainly positive and can be described as happy for most of the days, apart from Mondays and Sundays where he also feels sad due to uncertainty thoughts on his imminent retirement. Regarding the cognitive behaviour, he is rarely involved into cognitive tasks (such as reading a book, playing sudoku) due to the lack of free time. Usually, on Wednesdays and on Sundays he is used to read a book for 30-60min.

After the retirement day, he uses the COUCH system for four more weeks (until 3/11/2019). During this time, he spends most of his time at home with his family. Furthermore, he has more time to socialize with friends and get involved in cognitive tasks. It is worth mentioning that during the 5th week (7/10/2019 – 13/10/2019) there is a dramatic change on his behaviour, translated primarily into a physical, emotional and social activity drop. As a result, in the 6th week (14/10/2019 – 21/10/2019) an intervention from the 'Council of Coaches' virtual agents takes place. An overview of the dataset can be seen in the following figures.

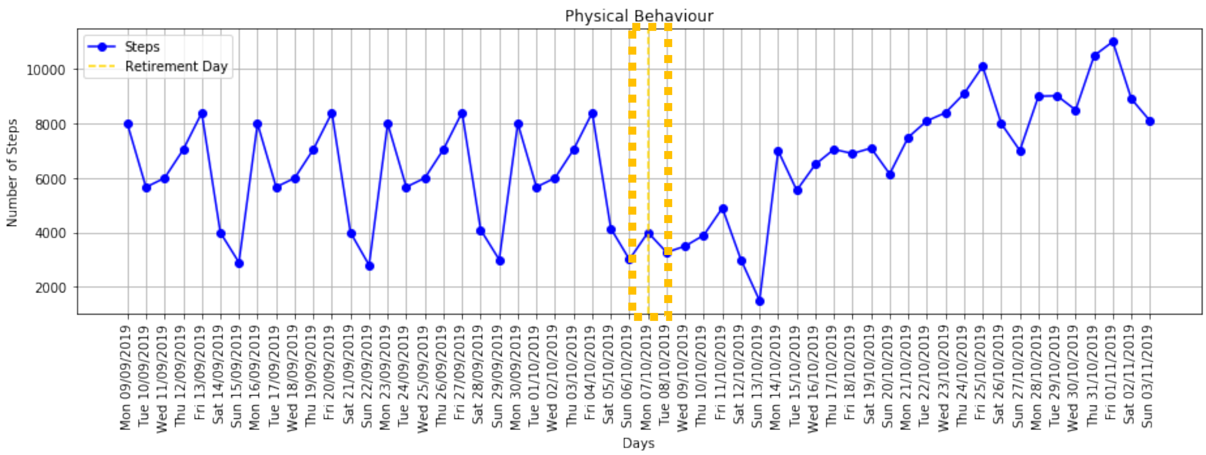

Figure 7: Overview of the physical behaviour time series (total number of steps per day).

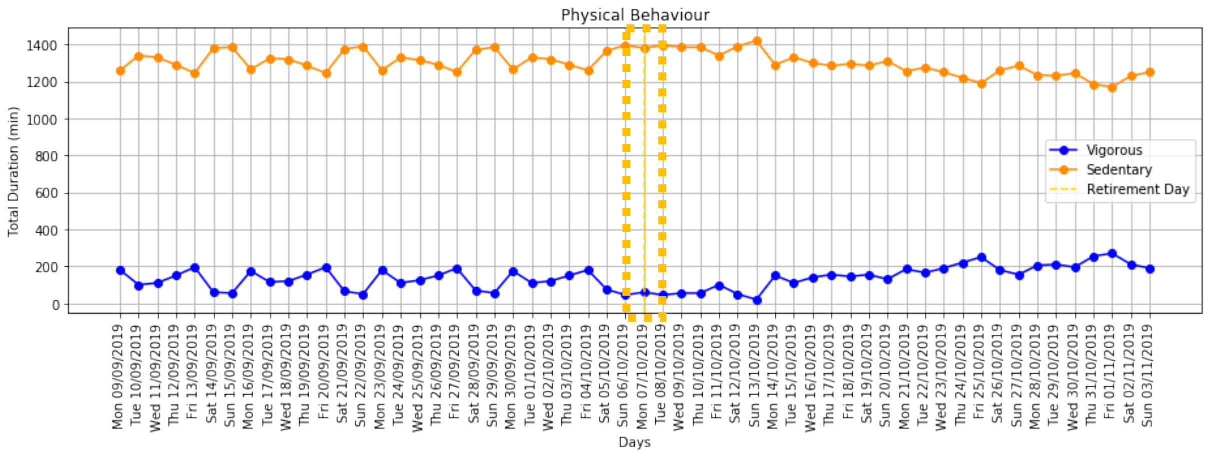

Figure 8: Overview of the physical behaviour time series (total duration of being sedentary or vigorously active per day).

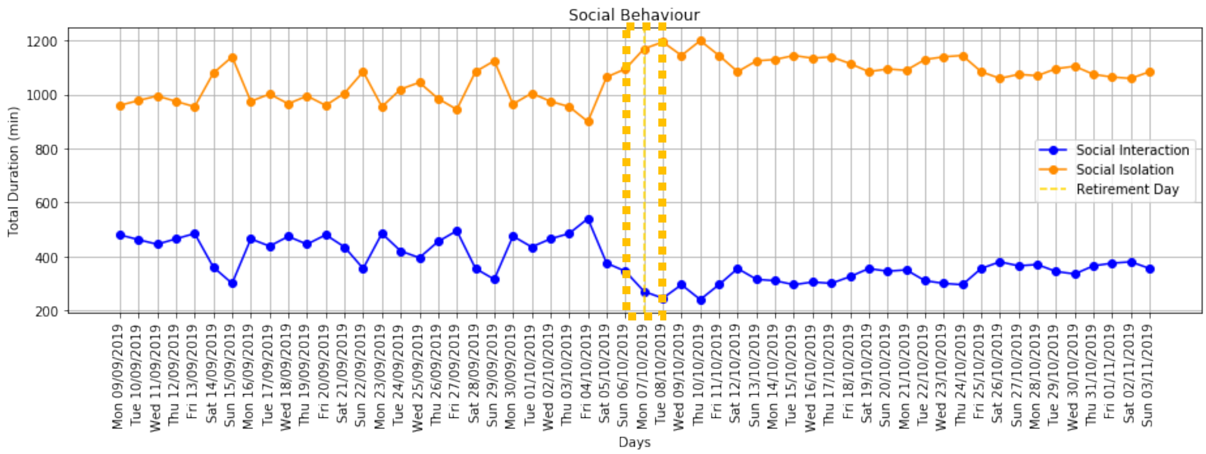

Figure 9: Overview of the social behaviour time series (total duration of being socially active or isolated per day).

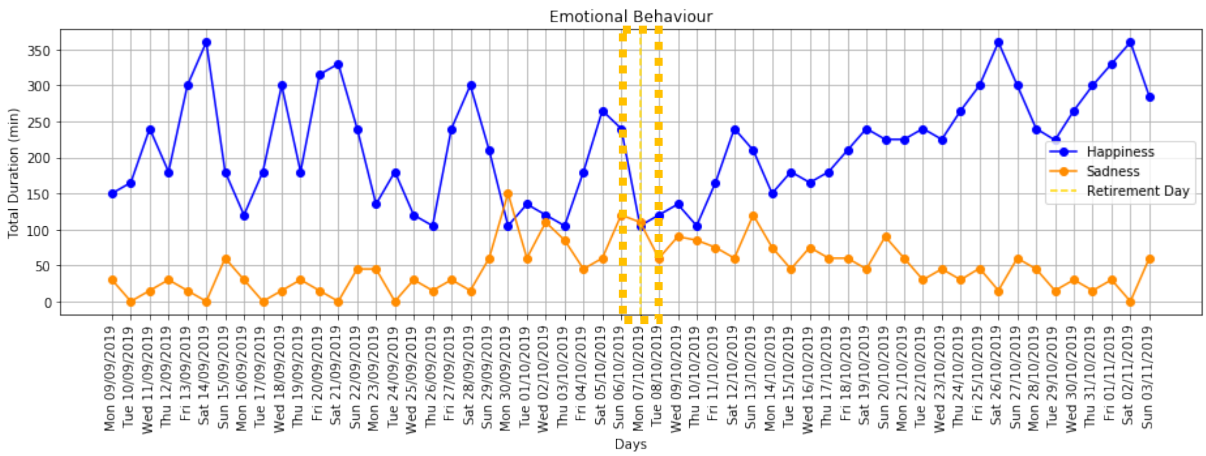

Figure 10: Overview of the emotional behaviour time series (total duration of being happy or sad per day).

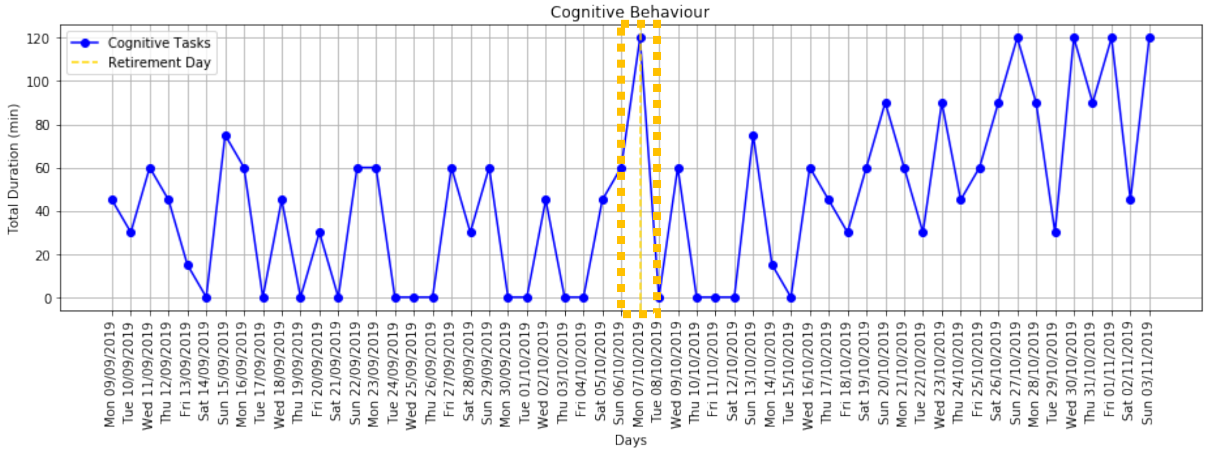

Figure 11: Overview of the cognitive behaviour time series (total duration of being involved in cognitive tasks per day).

5.1.2 Results

For practical reasons, and in order to compare the capabilities of each CPD method, all the methods are tested to detect the changes in the steps (short-term physical behaviour) time series (see Figure 12 to Figure 15). This time series is selected for this preliminary comparison since it includes periodicities and variations that may capture quite reasonably the dynamics of a realistic short-term behaviour pattern.

Alternating colours designate segments with estimated change points (the pink areas mark the estimated start and end point for a change, while dashed lines mark the time where a true change occurred). For the sake of clarity, the first dashed line (07/10/2019) represents the start of the retirement day, the second dashed line (14/10/2019) represents start date for the intervention of COUCH and the third dashed line (22/10/2019) represents the start of the next week after the intervention (when the user really starts following the recommendations of the coaches). Thus, we could say that the dashed lines represent the tipping points or starting dates of a change in the user behaviour.

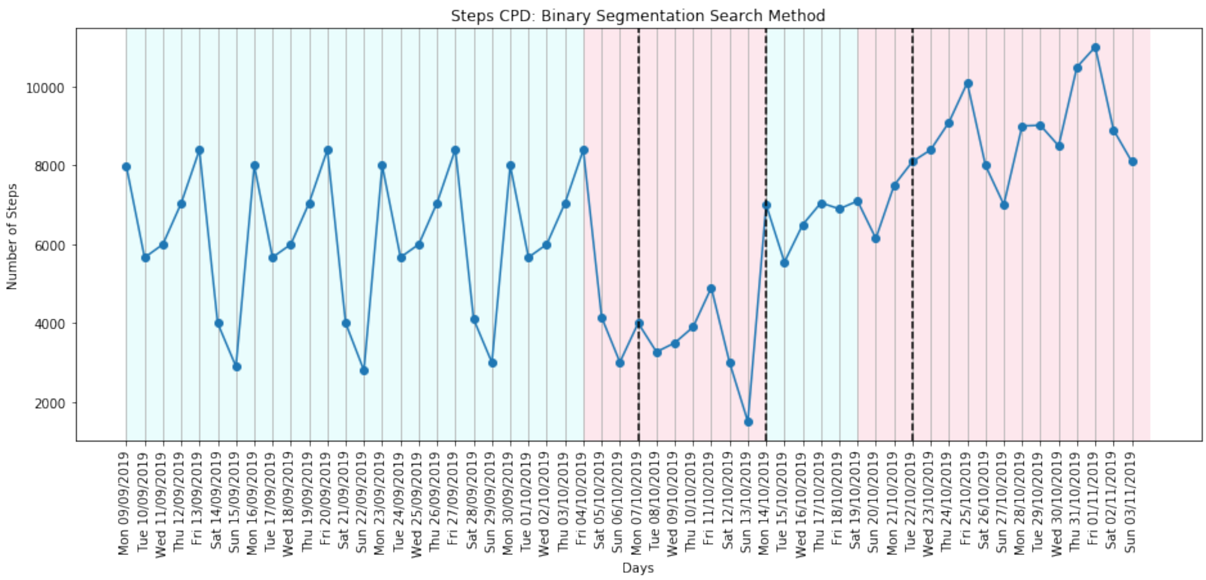

Figure 12: CPD for physical behaviour (steps) via Binary Segmentation Search Method.

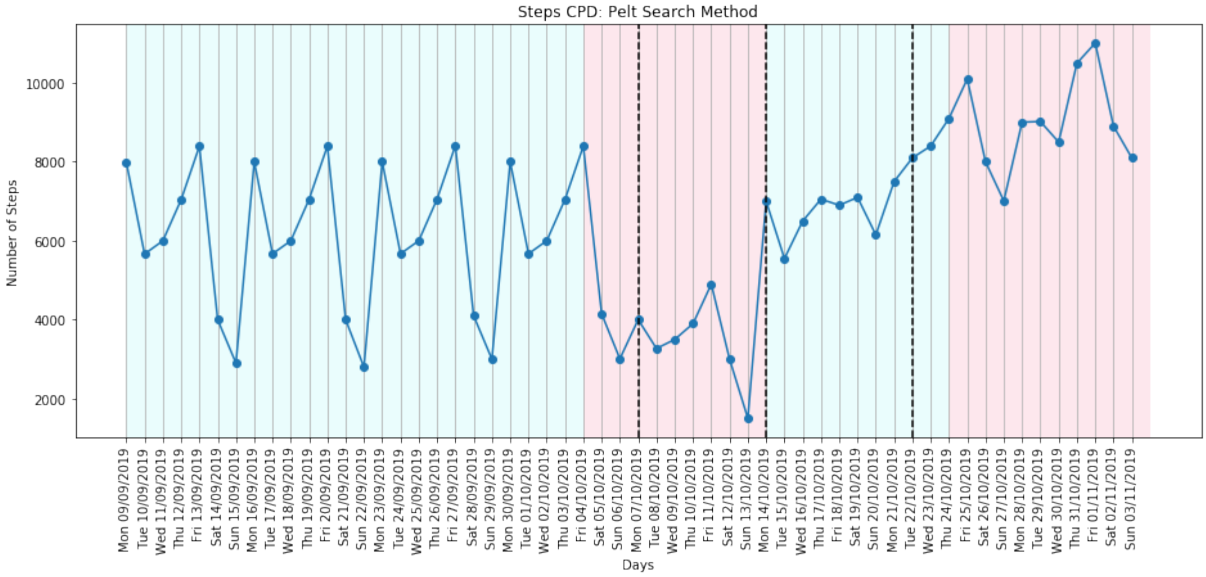

Figure 13: CPD for Physical Behaviour (steps) via Pelt Search Method.

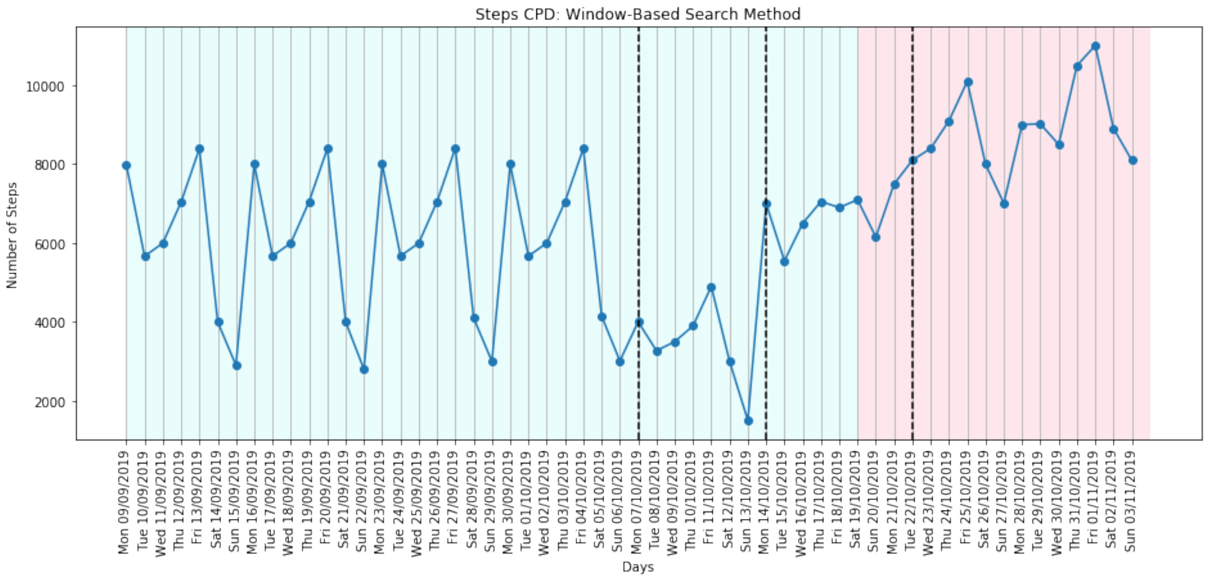

Figure 14: CPD for Physical Behaviour (steps) via Window-Based Search Method.

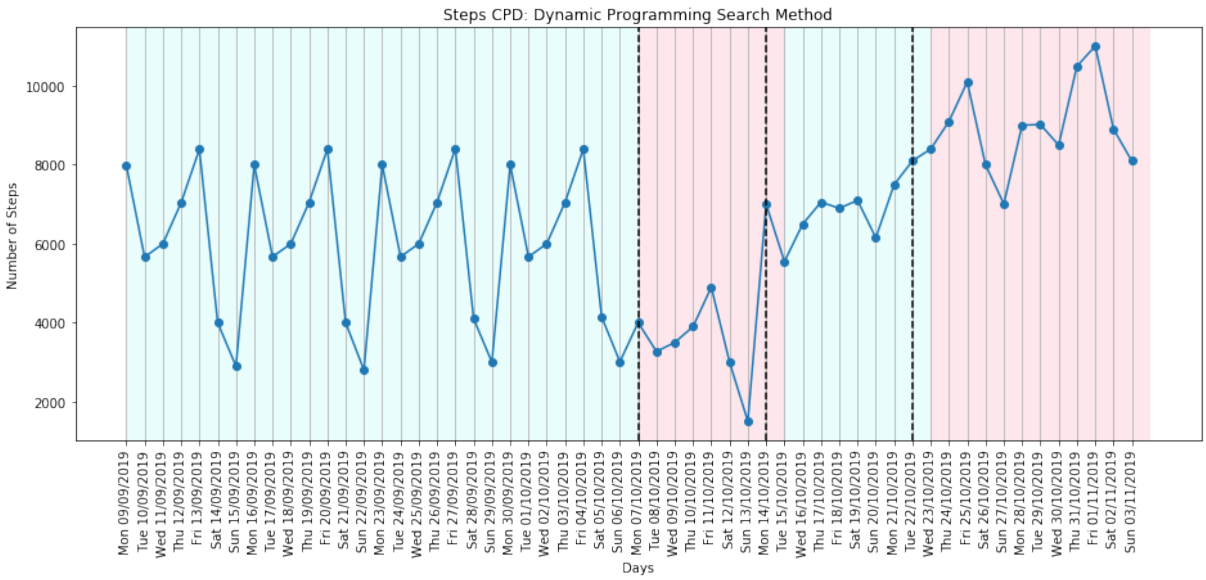

Figure 15: CPD for Physical Behaviour (steps) via Dynamic Programming CPD search method.

Table 2 shows the performance metrics computed for the four CPD search methods. From the results, it is clear that the Dynamic Programming search methods performs best among the considered ones, followed by the PELT. Thus, the Dynamic Programming search method is used hereafter for the evaluation of the rest of short-term behaviour time series.

Table 2: Evaluation of CPD methods for detecting change on steps (physical behaviour).

	Binary Segmentation	Pelt	Window-Based	Dynamic Programming
Precision / Recall	0.67 / 0.67	0.67 / 0.67	0.34 / 1.0	0.67 / 0.67
Rand index	0.90	0.92	0.73	0.98
Hausdorff	3.0	3.0	12.0	1.0

Figures 16 to 22 present the CPD results using the Dynamic Programming search method for detecting the changes on physical (see Figure 15, Figure 16 and Figure 17), social (see Figure 18 and Figure 19), emotional (see Figure 20 and Figure 21) and cognitive behaviours (see Figure 22). From the results it can be seen that the used method detects most closely the start and end of the emulated changes for those time series depicting clear transitions or variations (physical behaviour: steps, sedentary activities duration, vigorous activities duration), to a lesser extent for those with less abrupt transitions (social behaviour: interaction, isolation; emotional behaviour: happiness) and with a low accuracy for those series showing a rather steady pattern. This result is quite interesting since it does not necessarily mean that the method does not work well for some time series but that some short-term behaviour time series may not reflect well-enough a given change. Otherwise, a lifestyle change may be reflected in some behavioural components but not necessarily in all of them.

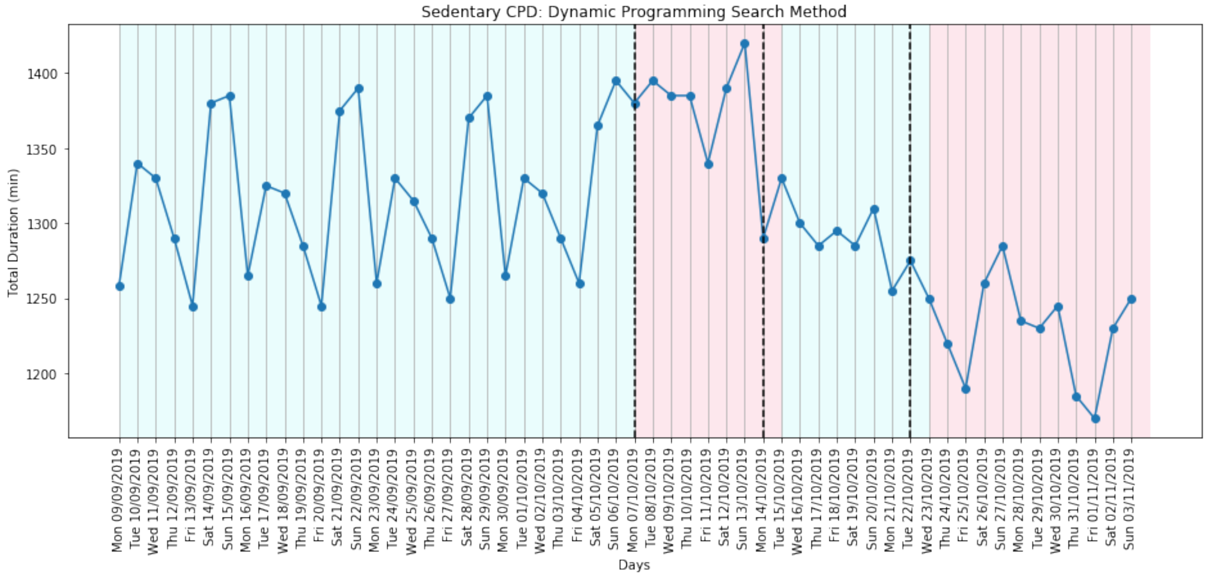

Figure 16: CPD for Physical Behaviour (sedentary duration) via Dynamic Programming CPD search method.

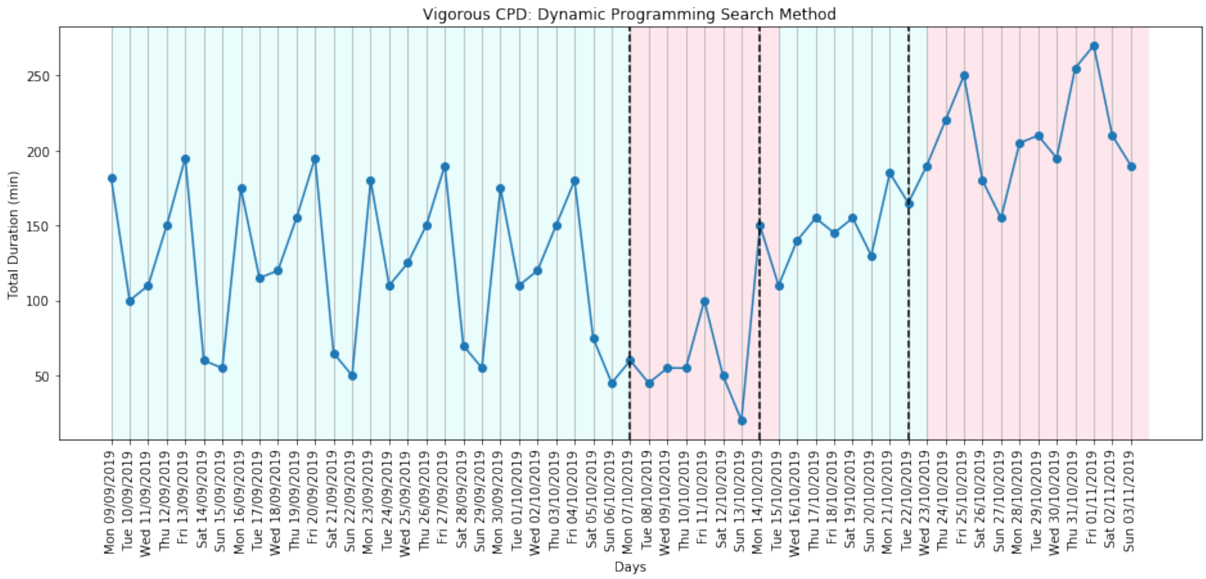

Figure 17: CPD for Physical Behaviour (vigorous duration) via Dynamic Programming CPD search method.

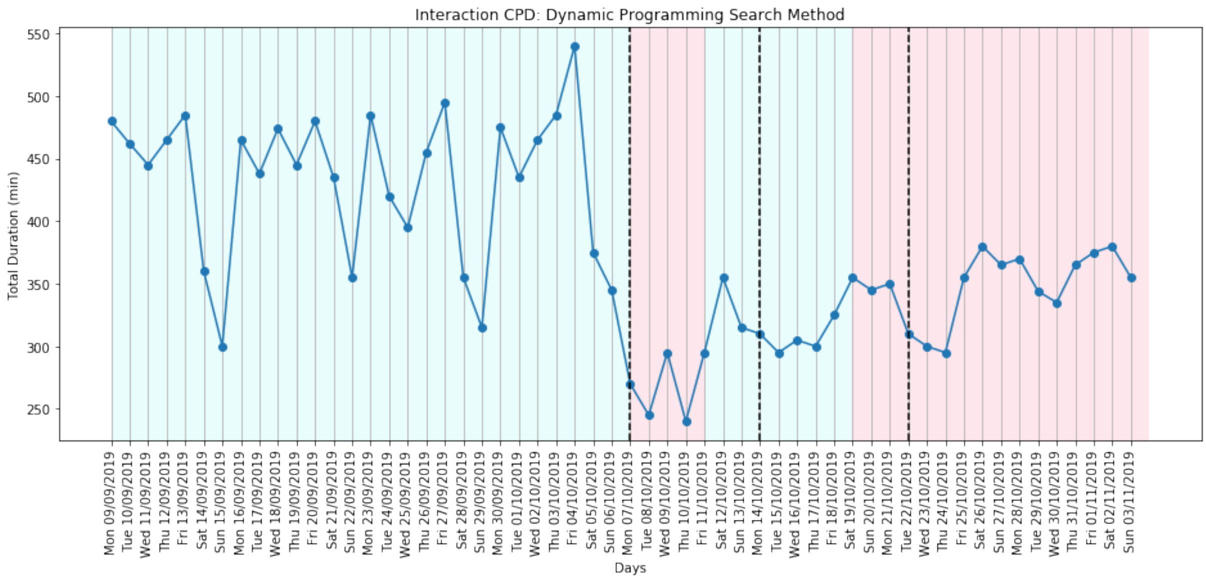

Figure 18: CPD for Social Behaviour (interaction duration) via Dynamic Programming CPD search method.

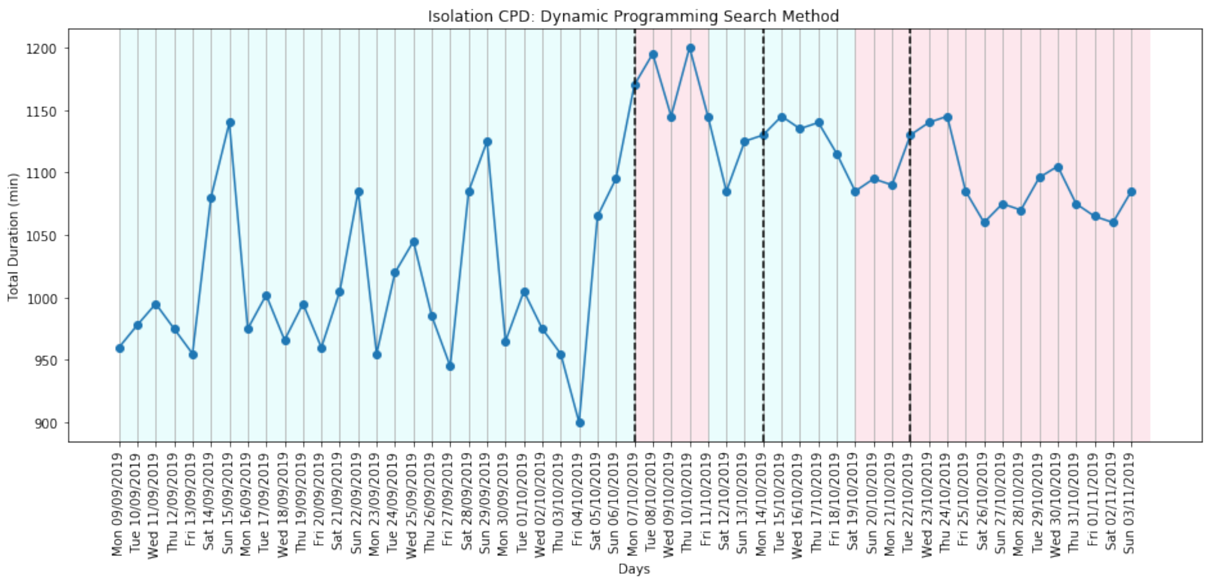

Figure 19: CPD for Social Behaviour (isolation duration) via Dynamic Programming CPD search method.

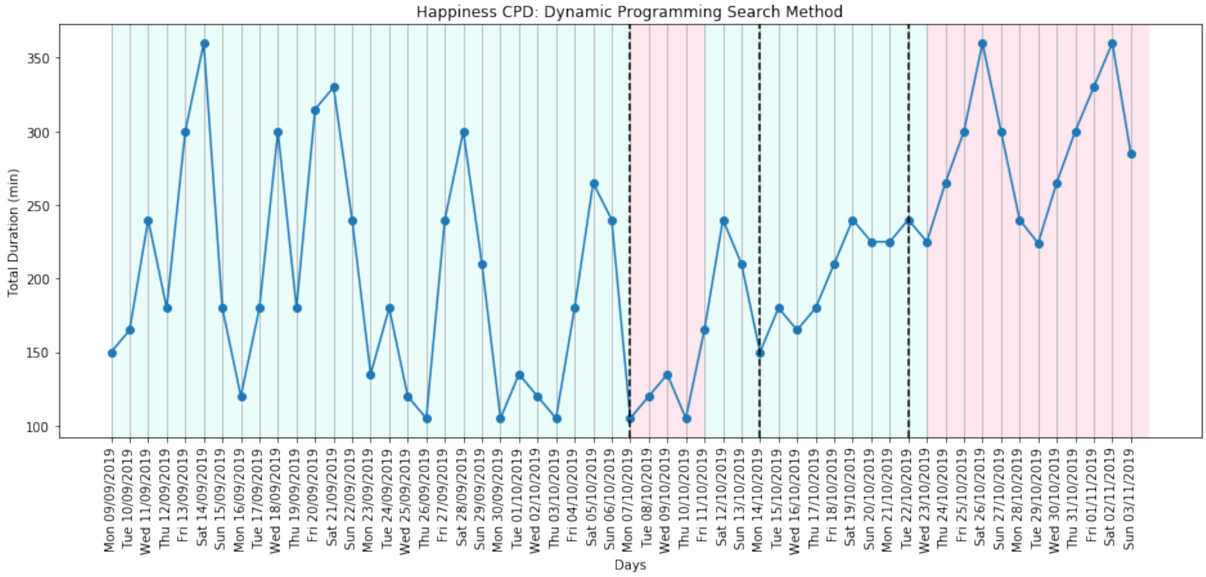

Figure 20: CPD for Emotional Behaviour (happiness duration) via Dynamic Programming CPD search method.

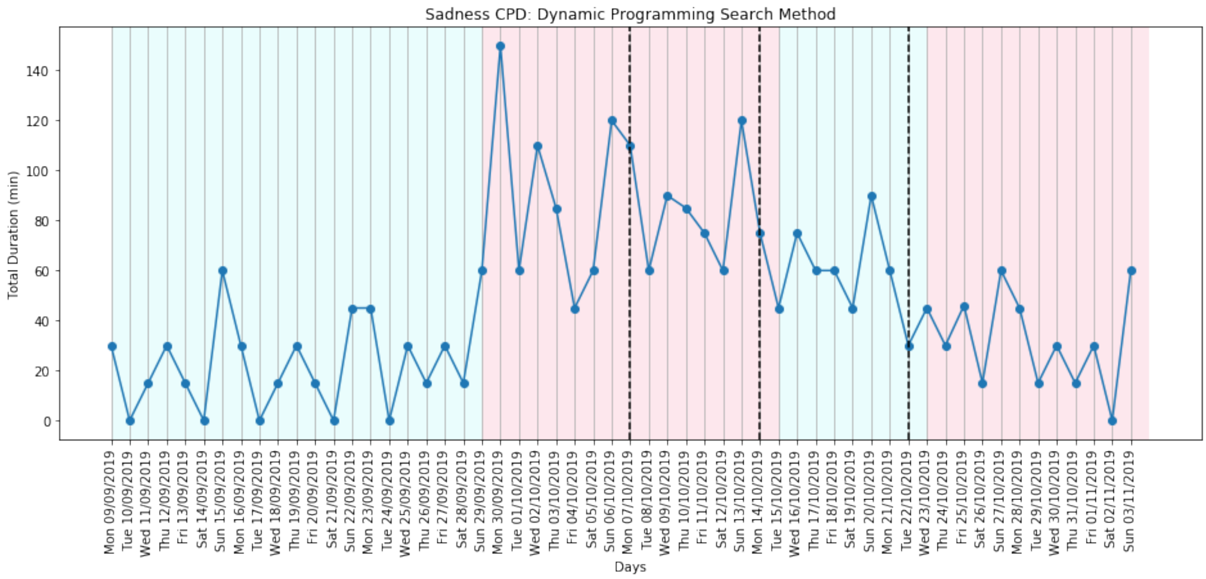

Figure 21: CPD for Emotional Behaviour (sadness duration) via Dynamic Programming CPD search method.

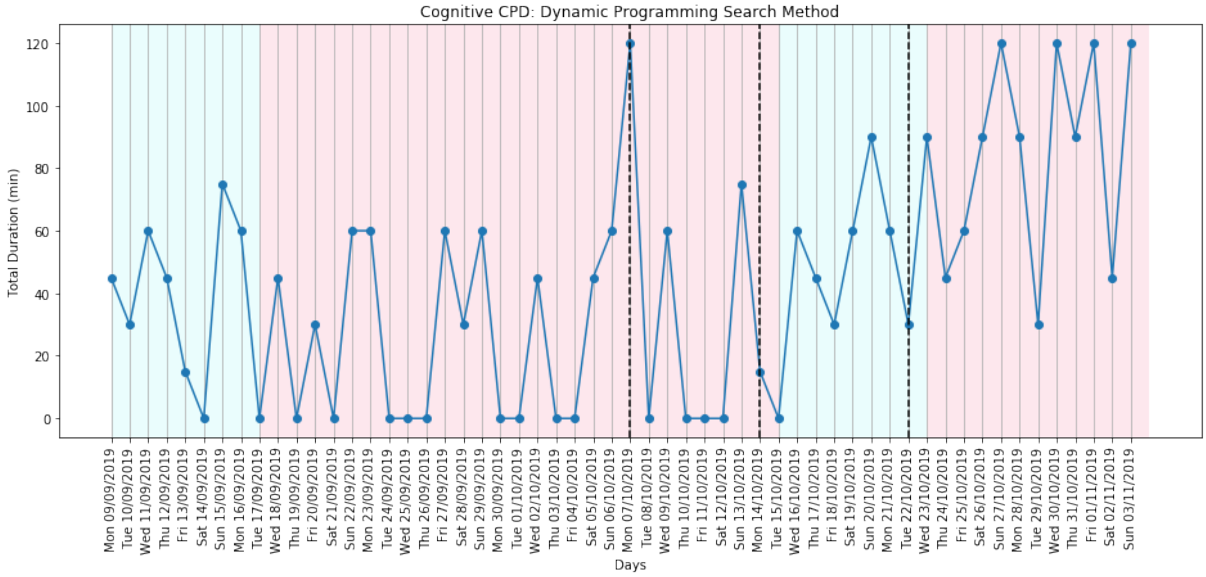

Figure 22: CPD for Cognitive Behaviour (total duration) via Dynamic Programming CPD search method.

Not only is important to detect the change but also estimate how relevant it is. Hence, in addition to the Dynamic Programming CPD search method, we decided to quantify the behaviour changes by comparing the percentage of change in a time series of elements. The percentage change evaluates the difference between the current and a prior element, and it is equal to the old value minus the new value, divided by the old value. In our case we are interested in computing the weekly percentage change, which accounts for periods of one week (e.g., how a weekday has differentiated compared to the same of the previous week).

In Figure 23, we can see that the percentage change for the physical behaviour (steps) does not deviate for the first four weeks. However, it changes dramatically right after the retirement day. It is worth mentioning that during the 5th week (7/10/2019 – 13/10/2019) there is a light percentage change (1st change), while during the 6th week (14/10/2019 – 21/10/2019), where an intervention takes place, there is a moderate change (2nd change). One week after the intervention and onwards, the user tends to follow a new lifestyle and maintain some new physical behaviour patterns (3rd change).

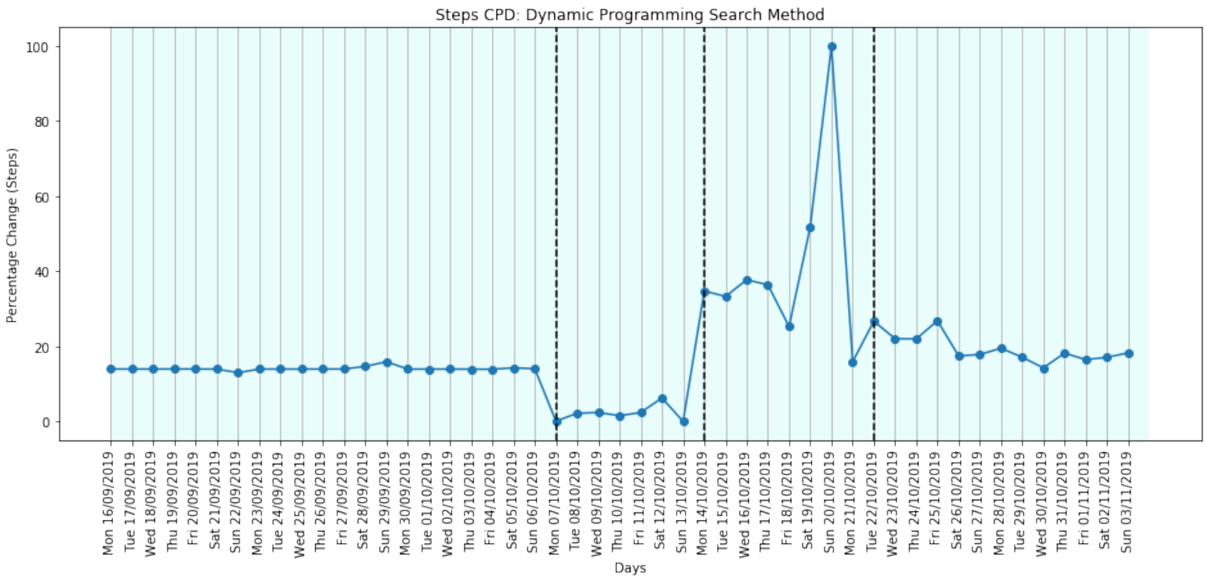

Figure 23: Percentage Change for Physical Behaviour (steps) via Dynamic Programming CPD search method.

Similarly, we can see the percentage change for the intensity of the physical behaviour (see Figure 24 and Figure 25), the duration for social behaviour (see Figure 26 and Figure 27), the emotional behaviour (see Figure 28 and Figure 29) and the cognitive behaviour (see Figure 30). The three changes can be seen in almost every behaviour, apart from the sadness and the cognition which do not reflect very much deviations on the user's emotional and cognitive behaviour.

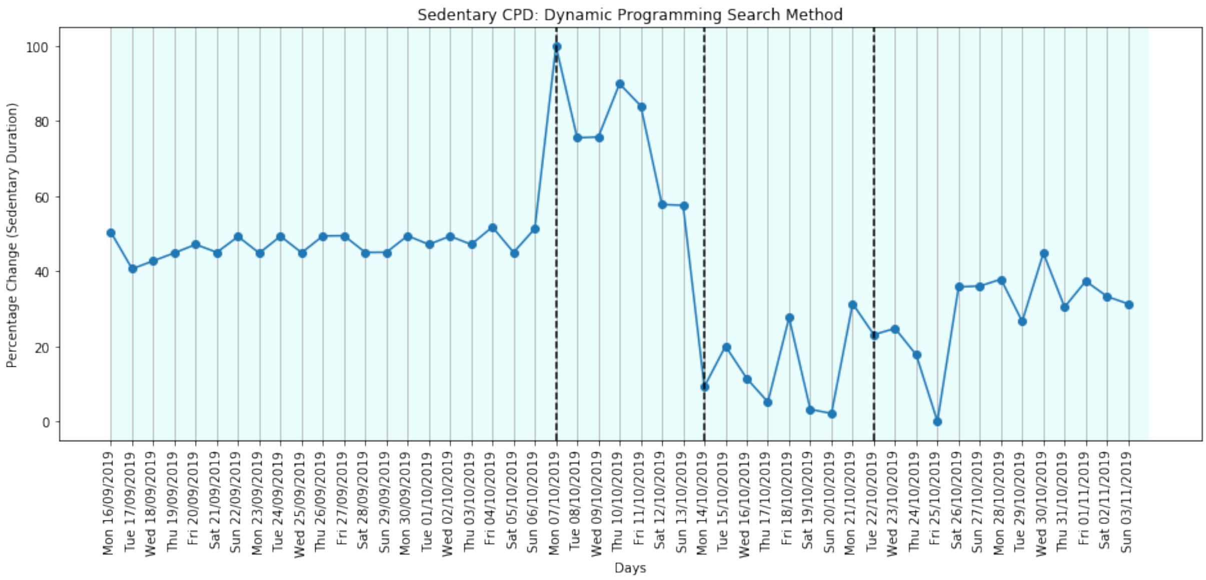

Figure 24: Percentage Change for Physical Behaviour (sedentary duration) via Dynamic Programming CPD search method.

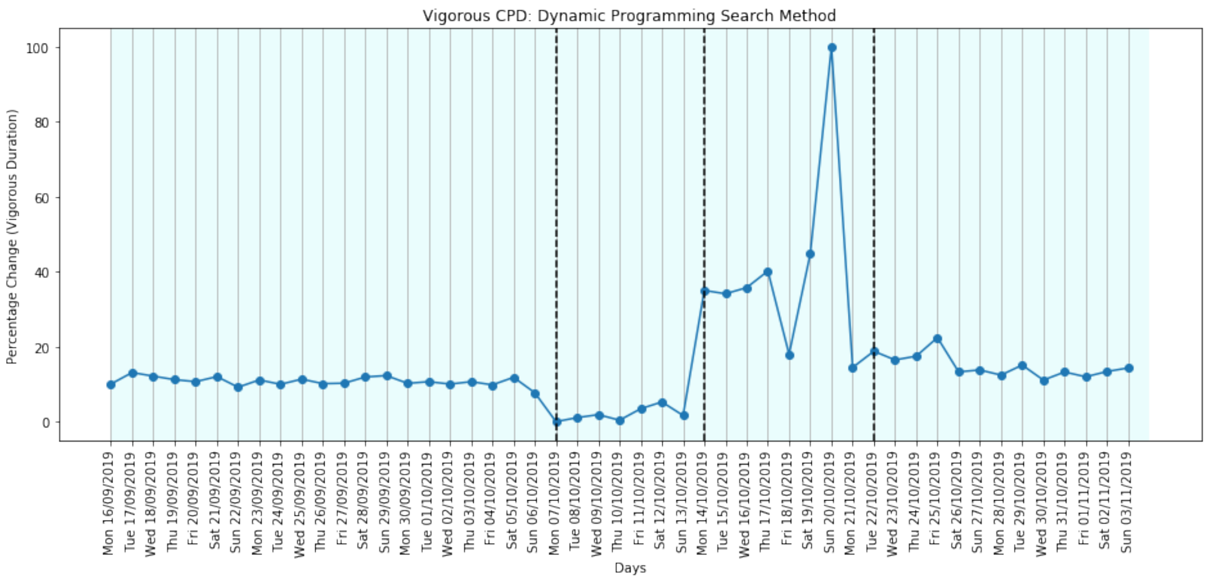

Figure 25: Percentage Change for Physical Behaviour (vigorous duration) via Dynamic Programming CPD search method.

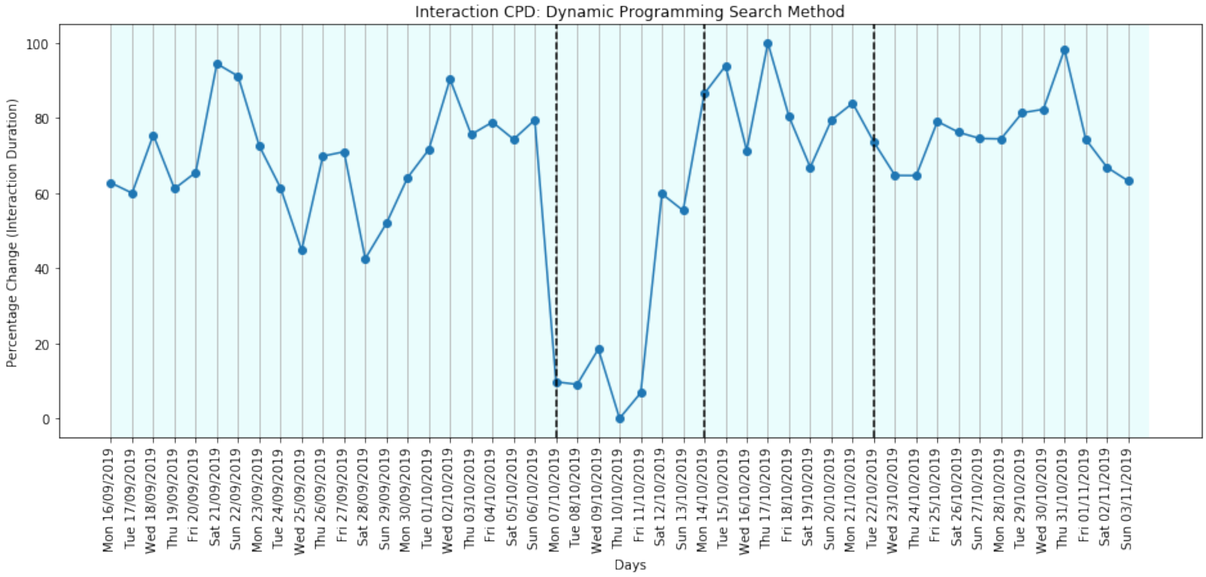

Figure 26: Percentage Change for Social Behaviour (interaction duration) via Dynamic Programming CPD search method.

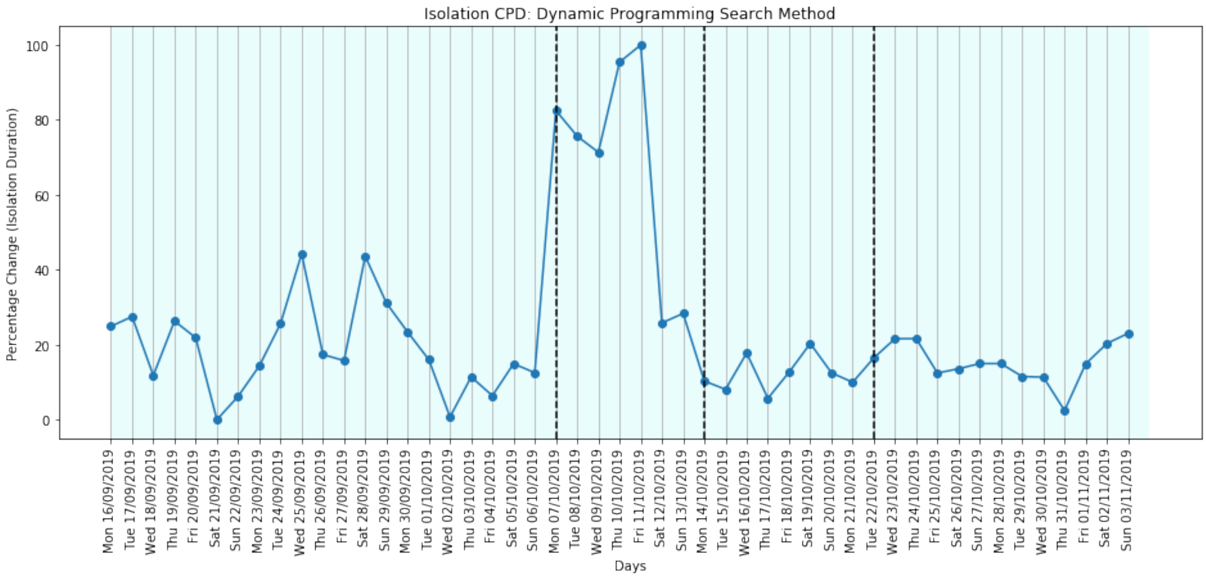

Figure 27: Percentage Change for Social Behaviour (isolation duration) via Dynamic Programming CPD search method.

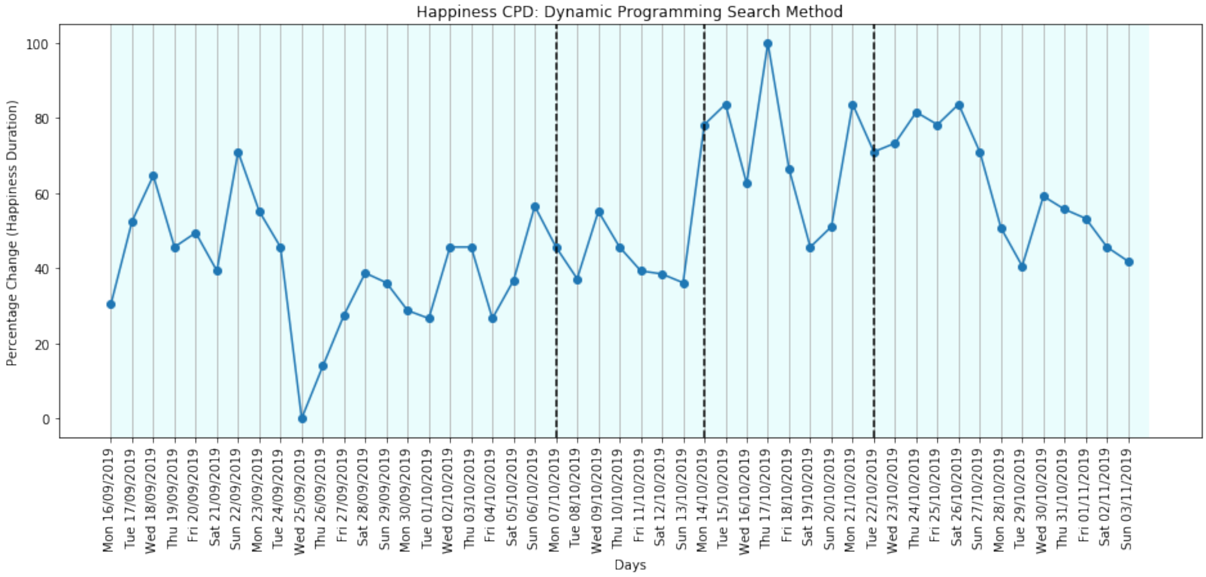

Figure 28: Percentage Change for Emotional Behaviour (happiness duration) via Dynamic Programming CPD search method.

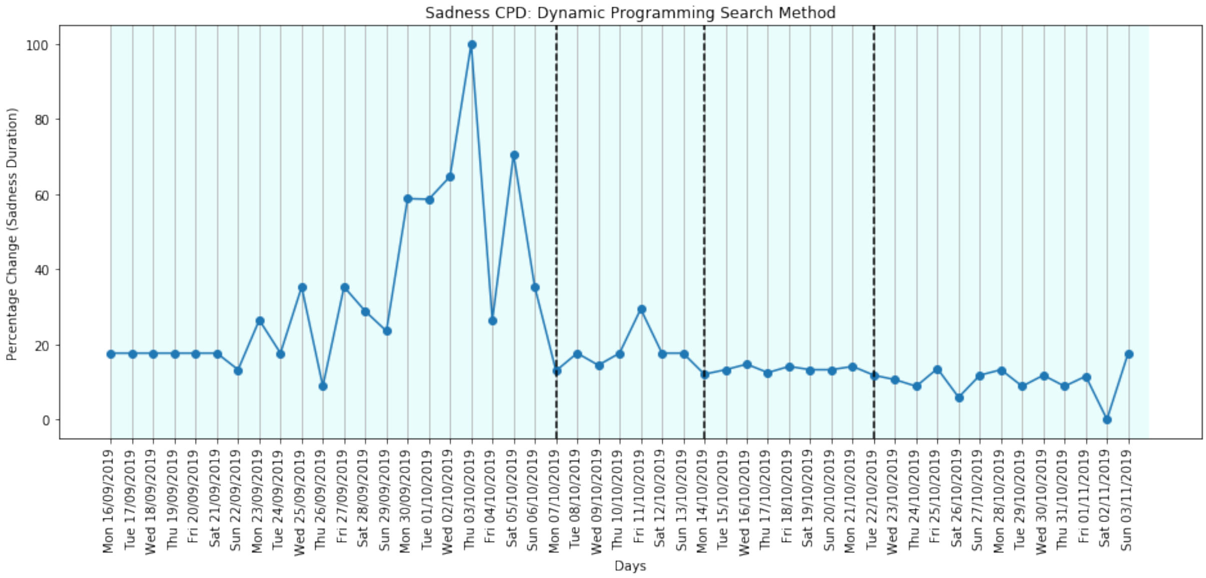

Figure 29: Percentage Change for Emotional Behaviour (sadness duration) via Dynamic Programming CPD search method.

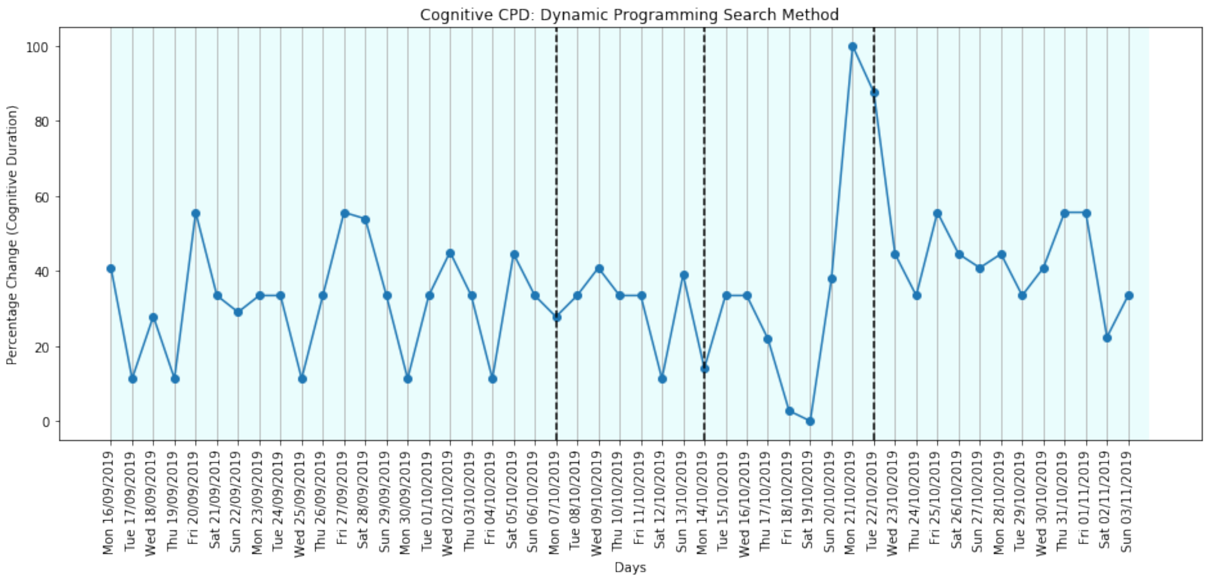

Figure 30: Percentage Change for Cognitive Behaviour (total duration) via Dynamic Programming CPD search method.

5.2 Experiment 2: "In the Wild" Dataset

This dataset represents data collected in an uncontrolled environment (also known as in the wild setting). This dataset consists of real data, acquired from a Fitbit device over a period of two months. Fitbit data represent physical behaviour, including the number of steps, the duration of sedentary and vigorous activities. Even though this dataset does not cover the whole spectrum of behaviours (i.e. behaviours such as social, emotional and cognitive are not examined here), our aim is to investigate the performance of our CPD approach for real-world data.

5.2.1 Experimental Setup

The Fitbit dataset consists of a single user's data for two months (from 22/5/2017 to 16/7/2017). We have arbitrarily chosen a period where the user was training daily to run a marathon. For the sake of comparison, we decided to investigate only two months similarly to the simulated dataset (Experiment 1). An overview of the Fitbit dataset can be seen in the Figure 31 and Figure 32.

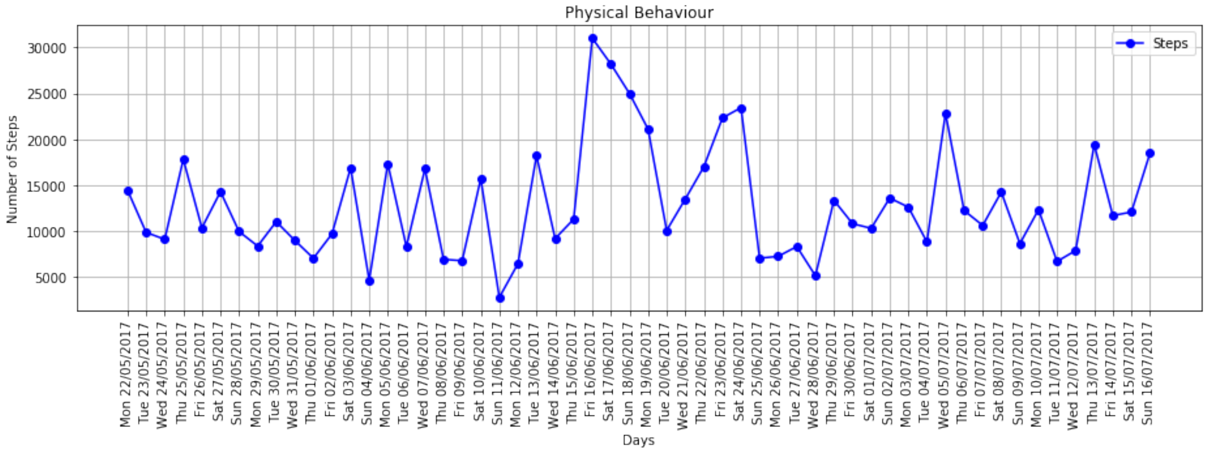

Figure 31: Overview of the physical behaviour (total number of steps per day) on Fitbit time series.

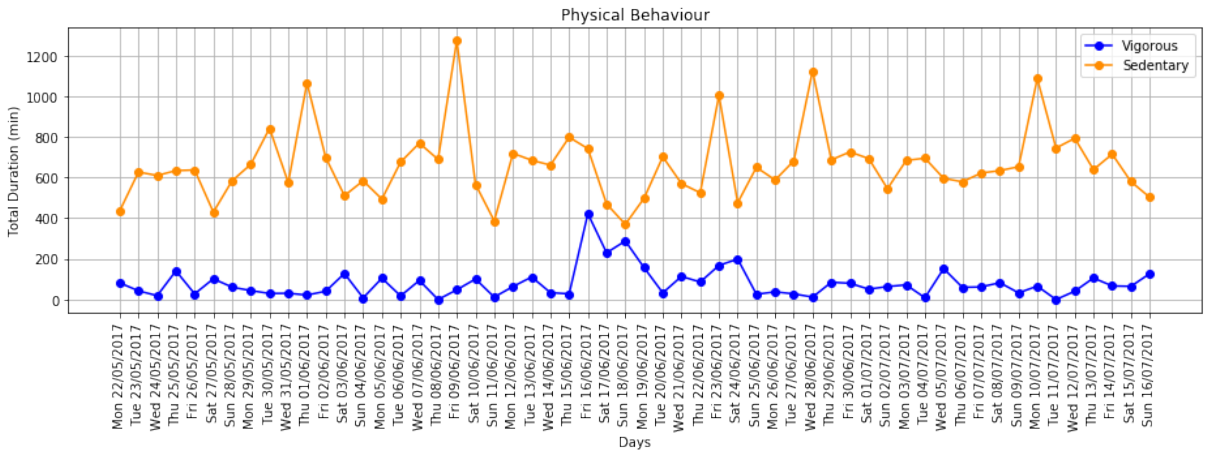

Figure 32: Overview of the physical behaviour (total duration of being sedentary or vigorously active per day) on Fitbit time series.

5.2.2 Results

In contrast to the controlled experiment, the Fitbit data were collected in the wild (uncontrolled experiment), and thus, the data as well as the behaviour changes are unlabelled. That means that the number of true changes is unknown and could differentiate significantly from the simulated dataset (where the number of true changes were at least three), since the behaviour changes can be affected by external factors such as seasonality and randomness. For comparability reasons, we have used the same CPD method used in the controlled experiment.

Overall, we can see that the user has been vigorously active at the end of the first month (exceeding 20,000 steps per day). According to Figure 33, Figure 34 and Figure 35, we can see that the CPD method reveals three main changes; the first one starts on Thursday 15/6, while the second starts on Sunday 25/6 and the third on Thursday 29/6.

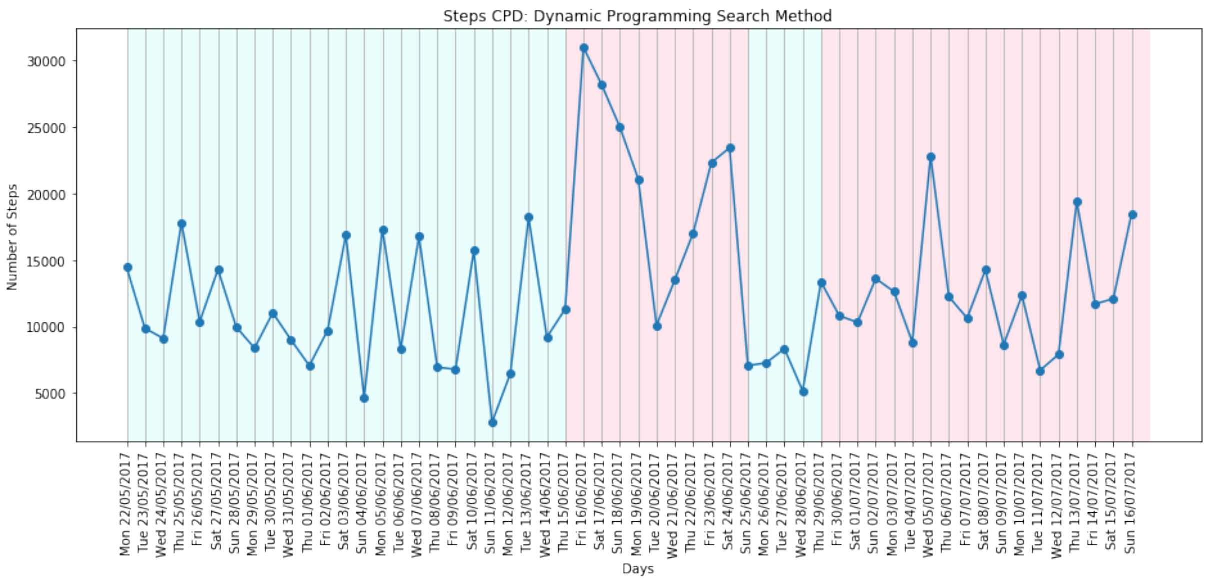

Figure 33: CPD for Fitbit Physical Behaviour (steps) via Dynamic Programming CPD search method.

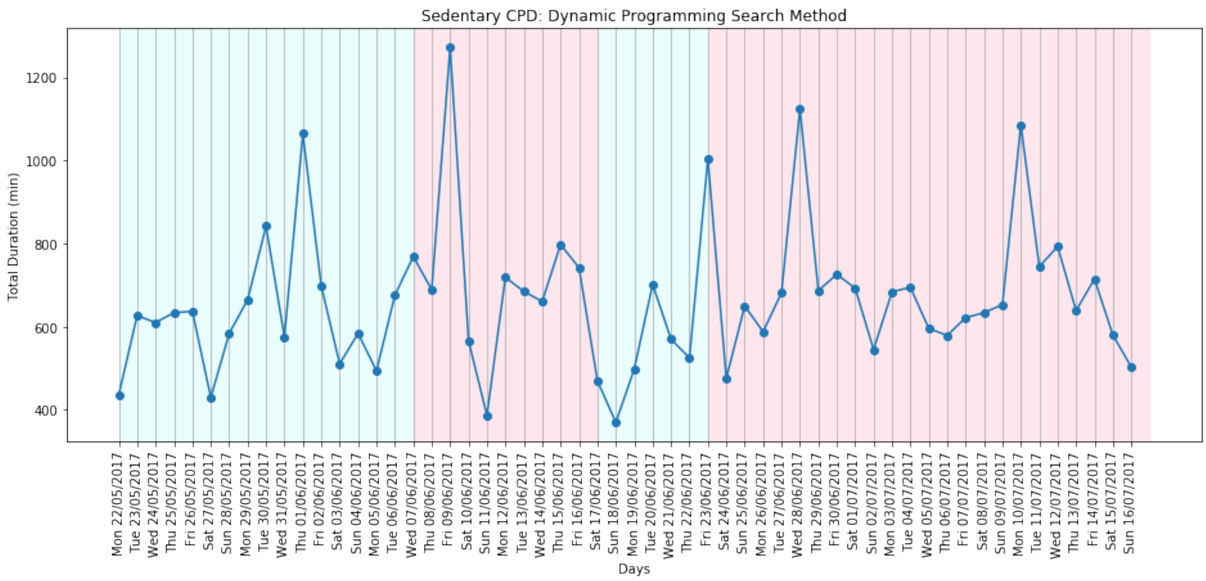

Figure 34: CPD for Fitbit Physical Behaviour (sedentary duration) via Dynamic Programming CPD search method.

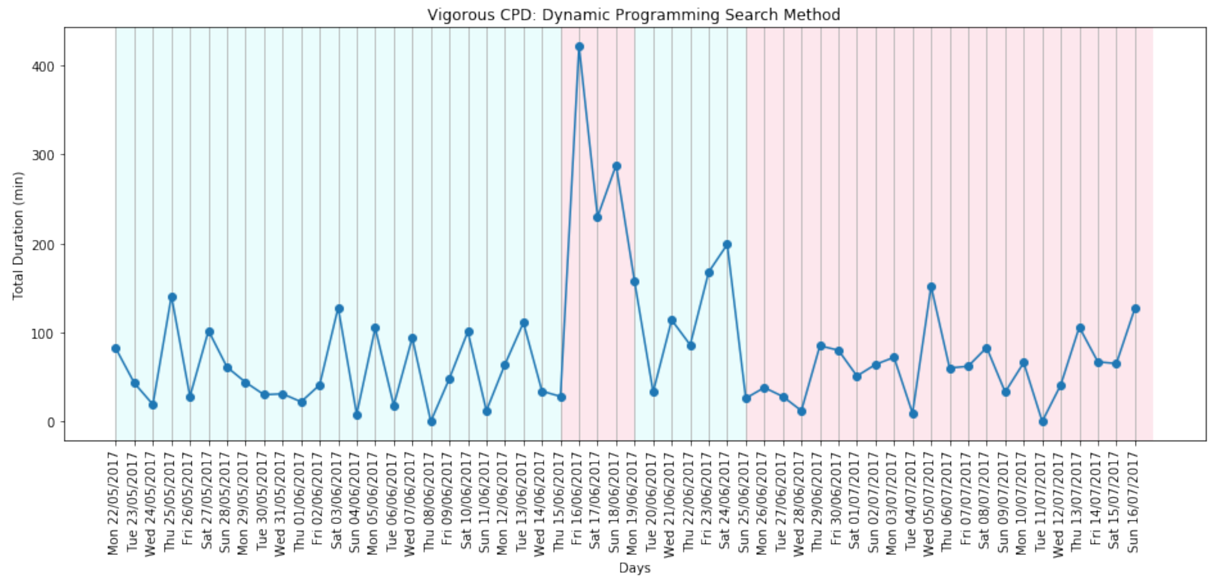

Figure 35: CPD for Fitbit Physical Behaviour (vigorous duration) via Dynamic Programming CPD search method.

In order to get some further insights on these changes, we quantify the behaviour changes by comparing the weekly percentage of change. In Figure 36, the biggest percentage change on steps occurs on Sunday 18/6. In Figure 37 the biggest change on sedentary duration is on Friday 9/6 and Wednesday 28/6. In Figure 38, there are more obvious deviations; the first big change starts on Sunday 18/6 and the second one on Wednesday 5/7. Specifically, on Friday 16/6 and on Sunday 18/6 there is a significant increase of the vigorous activities. Thus, we can estimate that at least two changes took place around Thursday 15/6 and Friday 30/6.

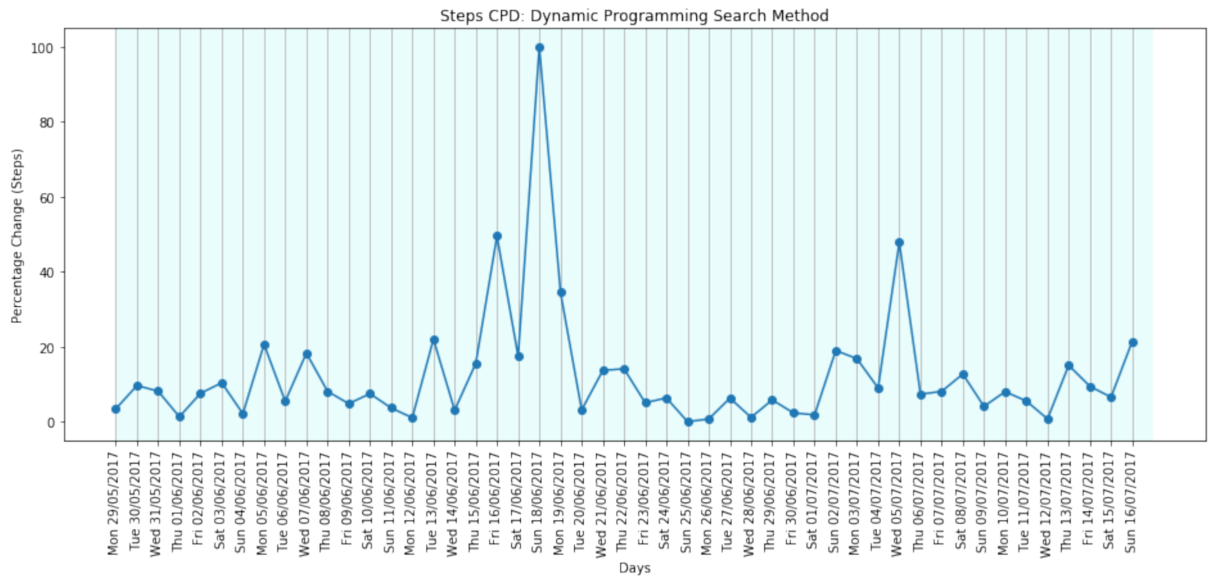

Figure 36: Percentage Change for Fitbit Physical Behaviour (steps) via Dynamic Programming CPD search method.

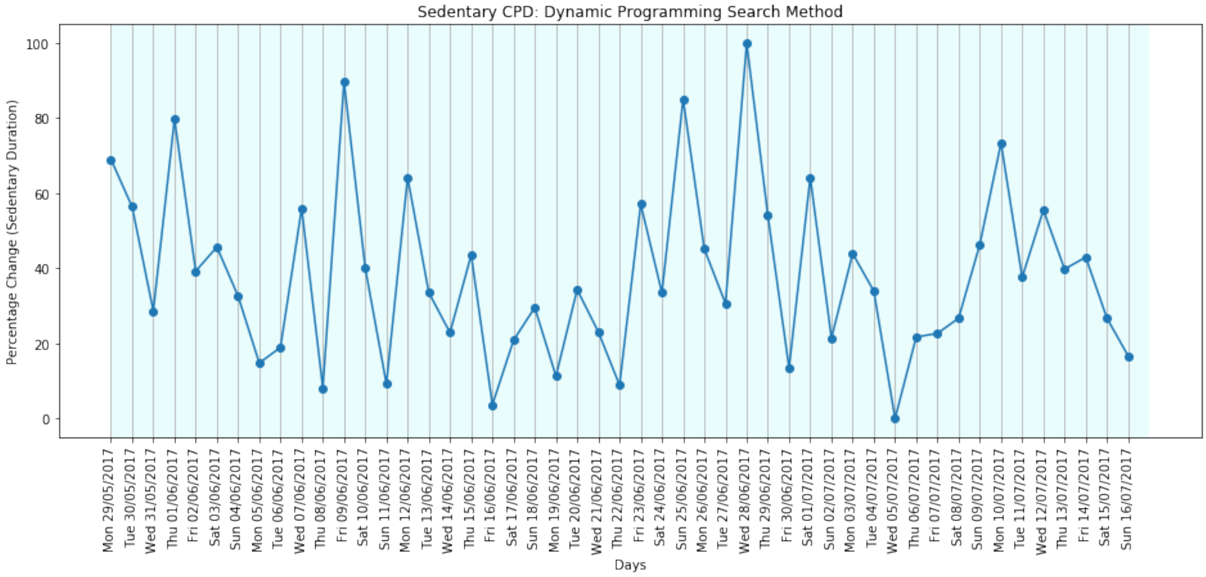

Figure 37: Percentage Change for Fitbit Physical Behaviour (sedentary duration) via Dynamic Programming CPD search method.

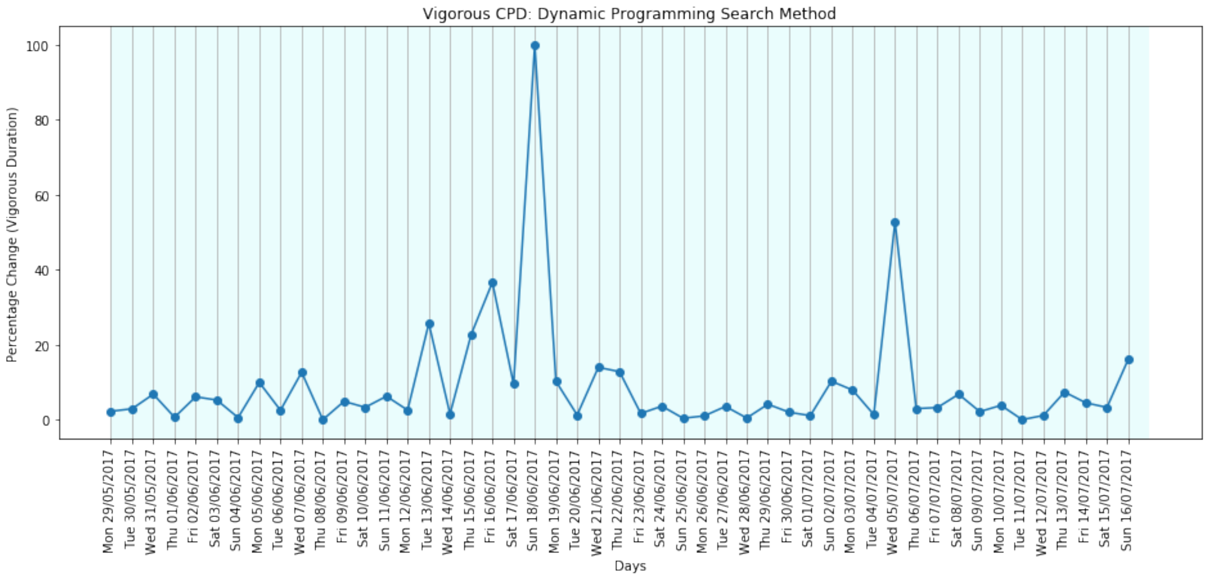

Figure 38: Percentage Change for Fitbit Physical Behaviour (vigorous duration) via Dynamic Programming CPD search method.

6 Discussion

6.1 Main Findings

This deliverable aims to present the proposed methodology for detecting changes between time periods (change detection) and determining the significance of the detected changes (change assessment). The methodology proposed after preliminary evaluation for the change detection is the Dynamic Programming CPD search method, while the change assessment is attained via the weekly percentage of change.

In order to validate our approach, we analysed data from two experiments. The first experiment consists of synthetic data with at least three relevant changes. The nature of the change in this case is the subject's retirement. In order to quantify the significance of the detected changes we calculated the percentage of change (variation) between one day and the same day of the previous week. Physical and social behaviours include more noticeable changes (compared to emotional and cognitive), and thus, it is easier to quantify their significance. Even though this approach seems reasonable, it needs further exploration in order to investigate the change detection on data with not significant deviations. Thus, a potential limitation of the current study is that some changes are not significant to be assessed. In order to ascertain whether this is a limitation of the method or the data a more thorough exploration must be performed in T4.4.

The second experiment consists of Fitbit data, collected in the wild, revealing information about user's physical behaviour. The nature of the change in this case is the training for a marathon and we can see that the user is being significantly active at the end of the first month. However, a limitation of this experiment is that the ground-truth for the detected changes is not known in advance. Similarly, as to the previous study, we quantify the significance of the detected changes by calculating the percentage of change (variation) between one day and the same day of last week. The detection method appears to spot quite precisely the most relevant changes shown in the time series. However, in order to interpret this appropriately, an accurate ground-truth labelling of when and why a behaviour change occurred is required. Here again, T4.4 is expected to serve well for this purpose.

6.2 Open Issues

Despite an actual evaluation of the proposed methods is not aimed at this deliverable, we considered it appropriate to prove our concept via two exemplary studies. We decided to use the Dynamic Programming CPD search method in order to detect when and for how long a change takes place. Although this technique showed promising detection results, it would be necessary to test it with more time series and possibly for longer time spans. Furthermore, in order to quantify these changes, we calculated the percentage change score. Even though the results seem to be promising for the synthetic dataset (1st experiment), we still need to validate our methodology using real data, where the behaviour changes are known. Our first attempt was to use a Fitbit dataset (2nd experiment), where we applied our methodology in real data in order to detect changes in physical behaviour. However, the fact that the number and the duration of real changes were unknown played a major role in the validation of our approach. The methods presented in sections 4 and 5 are intended to be further evaluated with the course of the planned future evaluations. This will take place during task T4.4.

7 Bibliography

Chatzopoulou, E. (2018, September 10). *Where do you start when designing for behaviour change?* Retrieved from UX Collective: https://uxdesign.cc/how-do-you-design-for-behaviour-change-790b9abefa08

Konsolakis, K., Hermens, H., Villalonga, C., Vollenbroek-Hutten, M., & Banos, O. (2018). Human Behaviour Analysis through Smartphones. *UCAMI 2018.* MDPI Proceedings 2018, 31, 1 .

Michie, S., van Stralen, M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science 6*, 42.

Oinas-Kukkonen, H. (2010). Behavior Change Support Systems: A Research Model and Agenda. *Persuasive Technology, 5th International Conference, PERSUASIVE 2010* (pp. 4-14). Copenhagen: Springer, Berlin, Heidelberg.

Pandas. (2019, September 21). *pct_change*. Retrieved from pandas.DataFrame: https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html

Paris-Saclay, E. (2018, January 3). *Ruptures - Search methods*. Retrieved from CNRS: http://ctruong.perso.math.cnrs.fr/ruptures-docs/build/html/detection/index.html

Prochaska, J. O., & DiClemente, C. C. (1982). Transtheoretical therapy: Toward a more integrative model of change. *Psychotherapy: Theory, Research & Practice*, 276-288.

Truong, C., Oudre, L., & Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing, Volume 167*.

Acknowledgements

The Council of Coaches 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.

Headings and titles in this document, as well as the Council of Coaches logo use the Comfortaa font, designed by Johan Aakerlund and Cyreal and licensed under the Open Font License[1].

Additional text in this document uses the Roboto font, designed by Christian Robertson and licensed under the Apache License, Version 2.0[2].

The Council of Coaches logo and Blobmen graphics were drawn freely in Inkscape, licensed under the GNU General Public License[3].

[1] Open Font License: http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&id=OFL_web

[2] Apache License, Version 2.0: http://www.apache.org/licenses/LICENSE-2.0

[3] Inkscape License Information: https://inkscape.org/about/license/