A FRAMEWORK FOR DESIGN AND PERSONALIZATION OF DIGITAL, JUST-IN-TIME, ADAPTIVE INTERVENTIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SUAT GÖNÜL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

NOVEMBER 2018

Approval of the thesis:

A FRAMEWORK FOR DESIGN AND PERSONALIZATION OF DIGITAL, JUST-IN-TIME, ADAPTIVE INTERVENTIONS

submitted by **SUAT GÖNÜL** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Halit Oğuztüzün
Head of Department, **Computer Engineering**

Prof. Dr. Ahmet Coşar
Supervisor, **Department of Computer Engineering, METU**

Examining Committee Members:

Prof. Dr. İsmail Hakkı Toroslu
Department of Computer Engineering, METU

Prof. Dr. Ahmet Coşar
Department of Computer Engineering, METU

Prof. Dr. Özgür Ulusoy
Department of Computer Engineering, Bilkent University

Prof. Dr. Pınar Karagöz
Department of Computer Engineering, METU

Assoc. Prof. Dr. Tansel Dökeroğlu
Department of Computer Engineering, TEDU

Date:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Suat Gönül

Signature :

ABSTRACT

A FRAMEWORK FOR DESIGN AND PERSONALIZATION OF DIGITAL, JUST-IN-TIME, ADAPTIVE INTERVENTIONS

Gönül, Suat
Ph.D., Department of Computer Engineering
Supervisor : Prof. Dr. Ahmet Coşar

November 2018, 150 pages

Adverse and suboptimal health behaviors and chronic diseases are responsible from a substantial majority of deaths globally. Studies show that personalized support programs yield better results in overcoming these undesired behaviors and diseases. Digital, just-in-time and adaptive interventions are mobile phone-based notifications that are being used to support people wherever and whenever needed in coping with the health problem. In this study, a framework is proposed for design and personalization of such interventions. The design part targets intervention designers and allows them to configure interventions that address specific needs of a particular health problem or population. The personalization part presents a reinforcement learning based mechanism to optimize intervention delivery strategies with respect to timing, frequency and type of interventions. Specifically, two reinforcement learning models, namely intervention-selection and opportune-moment-identification, are employed simultaneously. The models are fed with data obtained pertaining to people's long-term and momentary contexts. While the intervention-selection model adapts the intervention delivery with respect to type and frequency, the opportune-moment-identification model tries to find the most opportune moment to send interventions. Two improvements over the standard reinforcement algorithms are proposed to boost the learning performance. First, a customized version of eligibility traces is employed to reward past actions throughout the agent's trajectory in a selective manner. Second, the transfer learning method is utilized to reuse knowledge across multiple learning environments. The design and personalization modules of the proposed approach are validated individually. For the design part, it is shown that the proposed approach addresses the requirements of the intervention design specifications extracted from the extant literature. It is also shown that the design mechanism was utilized to design interventions for a real-life care program. The personalization part is validated mainly via a simulated case-study. Four personas are simulated with differentiating parameters in their daily activities, preferences on specific intervention types and attitude towards the targeted behavior. The results show that the improved algorithms yield better results compared to the standard versions and capture the simulation variations associated to the personas. A small-scale real-life case study has also been conducted utilizing a preliminary version of the proposed personalization method. Better results were obtained by the proposed approach compared to the base algorithm and a fixed intervention delivery schedule.

Keywords: just-in-time adaptive interventions, intervention scheduling optimization, reinforcement learning, m-health

ÖZ

DİJİTAL, ANLIK, UYARLANABİLİR MÜDAHALELERİN DİZAYN VE KİŞİSİLLEŞTİRİLMESİNE YÖNELİK BİR SİSTEM

Gönül, Suat
Doktora, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi : Prof. Dr. Ahmet Coşar

Kasım 2018 , 150 sayfa

Dünya çapındaki ölümlerin büyük çoğunluğunun sebebi sağlığa uygun olmayan davranışlar ve kronik hastalıklardır. Çalışmalar, kişiye özgü destek programlarıyla bu istenmeyen davranış ve hastalıkların üstesinden gelinmesinde daha iyi sonuçlar alındığını göstermektedir. Dijital, anlık ve uyarlanabilen müdahaleler herhangi bir konum ve zamanda, sağlık problemleriyle başetmek üzere uygulanan cep telefonu bildirimleridir. Bu çalışmada, sağlıkla ilgili müdahalelerin dizaynı ve kişiselleştirilmesine yönelik bir sistem önerilmektedir. Çalışmanın müdahale dizaynına yönelik kısmı müdahale tasarımcılarını hedeflemekte ve belirli bir sağlık problemi ya da kitlenin gereksinimlerine uygun müdahaleler tanımlamalarını sağlamaktadır. Kişiselleştirme kısmıysa müdahale uygulama stratejilerini müdahalenin zamanlamasına, sıklığına ve çeşidine göre eniyileyen pekiştirmeli öğrenme tabanlı bir mekanizma sunmaktadır. Bu mekanizma müdahale seçme ve uygun an tespitine yönelik, birbirleriyle senkronize bir şekilde çalışan, iki farklı pekiştirmeli öğrenme modeli içermektedir. Bu modeller kişilerin uzun-vadeli ve anlık bağlamlarıyla ilgili toplanan verilerle beslenmektedir. Müdahale seçme modeli müdahale uygulamasını sıklık ve çeside göre

uyarlarken, uygun an belirleme modeli uygulamanın zamanlamasına etki etmektedir. Öğrenme performansını arttırmak amacıyla, standart algoritmaların üzerine iki iyileştirme önerilmektedir. İlk olarak, öğrenme vekilinin takip ettiği yol boyunca geçmiş adımlarını seçici bir şekilde ödüllendirmesine yönelik uyarlanmış uygunluk izleri önerilmektedir. İkinci iyileştirmedeyse öğrenme transferi yöntemi belirli bir pekiştirmeli öğrenme ortamında öğrenilen bir bilginin birden ortamda yeniden kullanılmasını sağlamaktadır. Dizayn ve kişiselleştirme modülleri ayrı ayrı doğrulanmaktadır. Dizayn kabiliyetleri, önerilen yaklaşımın literatürden elde edilen müdahale gereksinimlerinin karşıladığını göstererek doğrulanmaktadır. Bunun yanı sıra dizayn mekanizması bir gerçek hayat tedavi programı için gerekli müdahaleleri tanımlamak için kullanılmıştır. Önerilen kişiselleştirme yöntemi ise esasen temsili deney yöntemiyle doğrulanmıştır. Günlük aktivitelerinde, belirli müdahale yöntemleriyle ilgili tercihlerinde ve hedeflenen davranış değişikliğine karşı tavırlarında farklılaşan dört farklı karakterin simülasyonu yapılmıştır. Sonuçlar iyileştirilmiş algoritmaların standart algoritmalara göre daha iyi sonuç verdiğini ve simülasyonu yapılan karakterlere atfedilen farklılıklara uygun şekilde tepki verdiğini göstermiştir. Önerilen kişiselleştirme yöntemiyle ilgili, ilk versiyonlarından biriyle yapılan küçük çaplı gerçek hayat deneyine yönelik sonuçlar da sunulmaktadır. Bu deneyde, önerilen yöntemle standard algoritmalara ve sabit müdahale uygulama planlarına göre daha iyi sonuçlar elde edilmiştir.

Anahtar Kelimeler: anlık uyarlanabilir bildirimler, bildirim planlama optimizasyonu, pekiştirmeli öğrenme, m-sağlık

To my beloved Elif Irmak and Gülseren

ACKNOWLEDGMENTS

I would like to express my candid thanks and appreciation to my supervisor Prof. Dr. Ahmet Coşar, for accepting me as his student. I would like to express my gratitude to Prof. Dr. İsmail Hakkı Toroslu, for his guidance and support throughout the study.

I am deeply grateful to Tuncay Namlı, without whose guidance and invaluable contributions, this work could not have been accomplished. I am highly thankful to Ali Anıl Sınacı and Mert Başkaya and all the other colleagues at SRDC Ltd., for their invaluable support throughout this study.

My biggest gratitude is to my dear wife Gülseren for her friendship, encouragement and for undertaking the great responsibility of growing our daughter when I was working for this study. I would have never had the strength to complete this work without her continuous support.

I also thank to my dear family for supporting me all the time.

I would also thank the Scientific and Technological Research Council of Turkey (TÜBITAK) for providing the financial means throughout this study.The research leading to these results has received funding also from the European Community's H2020 Program (H2020) under grant agreement no ICT-689444, POWER2DM Project (Predictive model-based decision support for diabetes patient empowerment).

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENTS xi

LIST OF TABLES xvi

LIST OF FIGURES xviii

LIST OF ABBREVIATIONS xx

CHAPTERS

1 INTRODUCTION 1

1.1 Chronic Disease Care Process Leading to Personalized Self-Management 1

1.2 Personalized Self-Management Support 2

1.3 A Framework for Design and Personalization of JITAIs . . . 3

1.4 Reinforcement Learning for Personalization of JITAIs 5

1.5 Outline 6

2 RELATED WORK 9

2.1 Personalized Support for Self-Management / Behavior Change 9

2.2 Computational Approaches Aiming at Personalization of Mobile Phone Notifications 11

2.3 Rewarding Past Actions 13

2.4 Transfer Learning Strategies in RL 14

2.5 System Design 15

3 BACKGROUND AND PRELIMINARIES 17

3.1 Self-Management of Health Problems 17

3.1.1 Action Plans 17

3.1.2 JITAI Components 18

3.2 Reinforcement Learning 20

3.2.1 Eligibility Traces 21

3.3 Transfer Learning 22

3.4 Social-Psychological Model of Prospective Memory and Habit Development 23

4 JITAI DESIGN 25

4.1 Conceptual Foundations of Self-Management 25

4.2 Template-Based JITAI Design 27

4.3 Rule Definition Language 33

5 THE JITAI PERSONALIZATION ALGORITHM 37

5.1 Alignment of Self-Management Concepts with RL 39

5.1.1 Analogy of Intervention Delivery and RL 39

Common characteristics in both analogies: 40

Specifics of the intervention-selection - RL analogy: 41

Specifics of the opportune-moment-identification - RL analogy: 41

5.1.2 Distributing the JITAI Components over the Two RL Models 41

5.2 RL Models 43

5.2.1 The Opportune-Moment-Identification Model . . . 44

5.2.1.1 Selective Eligibility Traces 46

5.2.1.2 Transfer Learning Across Opportune-Moment-Identification Environments . 48

Definition 1: (Common Policy - CP): 50

Definition 2 (State Classifier - SC): . 50

5.2.2 The Intervention-Selection Model 51

5.3 Overall Algorithm 58

6 COMMUNICATION ENGINE: THE SOFTWARE REALIZING THE JITAI DESIGN AND DELIVERY PLANNING 63

6.1 Reactive Programming within Lambda Architecture 63

6.2 Component Architecture 65

7 VALIDATION 69

7.1 Validating the JITAI Design Mechanism 69

7.2 Simulated Case Study 73

7.2.1 An Imaginary Action Plan 73

7.2.2 An Imaginary Set of Interventions Identified for the Targeted Behavior 74

7.2.3 Persona Simulation 75

7.2.3.1 Configuring the Habit Formation Model 75

7.2.3.2 Simulating Daily Activities 76

7.2.3.3 Simulating Reactions to Interventions 78

7.2.3.4 Simulating Behavior Performance . . 78
7.2.4 Hypotheses 79
7.3 Results 80
7.3.1 Validation of Hypotheses 80
7.3.2 Improvements on the Base Algorithm 85
7.3.2.1 Results for the Intervention-Selection Model 85
Complexity of the Model: 87
7.3.2.2 Results for the Opportune-Moment-Identification Model 88
Complexity of the Model: 90
7.3.3 Real-Life Case Study 90
7.3.4 Validation of the Reward Function Instantiations . 92
7.3.5 Validation of the Opportune-Moment-Identification Model Reward Function 92
7.3.6 Validation of the Intervention-Selection Model Reward Function 95
8 DISCUSSION 97
9 CONCLUSION 101
REFERENCES 103
APPENDICES
A BUILT-IN CONSTRUCTS OF RULE DEFINITION LANGUAGE . . 117
A.1 Tailoring Variables 117
A.2 Temporals 118

A.3 Placeholders 118

A.4 Temporal Index 120

A.5 Logical Operators 121

A.6 Behavior Change Techniques 121

B ANALYSIS OF CALO-RE TAXONOMY 123

C EXAMPLE INTERVENTION DEFINITIONS DRIVEN BY THE CALO-RE TAXONOMY 127

D EXAMPLE INTERVENTION DEFINITIONS DRIVEN BY THE RESOURCES AVAILABLE IN LITERATURE 131

E EXAMPLE INTERVENTION DEFINITIONS FROM THE POWER2DM REAL-WORLD CASE STUDY 135

F INTERVENTION DEFINITIONS FOR THE SIMULATED CASE STUDY 143

CURRICULUM VITAE 147

LIST OF TABLES

TABLES

Table 4.1 Elements of JITAI design template 31

Table 7.1 Elements of a daily activity 77

Table 7.2 Preferences associated to simulated personas on intervention types . 78

Table 7.3 Real-world experiment results 91

Table B.1 Implementation feasibility for the BCTs of CALO-RE taxonomy . . 123

Table C.1 Intervention example implementing the *setting grated tasks* technique 127

Table C.2 Intervention example implementing the *relapse prevention/coping planning* technique 128

Table D.1 Intervention example encouring a person to take a break after 30 minutes of prolonged sitting 131

Table D.2 Intervention example adjusting the insulin dosage based on the Predictive 303 algorithm 132

Table E.1 Intervention example for blood glucose monitoring behavior 135

Table E.2 Intervention example for carbohydrate monitoring behavior 137

Table E.3 Intervention example for physical exercise behavior 138

Table E.4 Intervention example for medication adherence behavior 140

Table F.1 Ordinary reminder intervention example 143

Table F.2 Reminder intervention example using social comparison BCT 144

Table F.3 Motivation intervention example praising the performance 145

LIST OF FIGURES

FIGURES

Figure 3.1 Partitioning of the daily time frame according to the action plans . . 18

Figure 4.1 Inter-relationship of self-management concepts 26

Figure 4.2 Relationship between interventions and decision points 29

Figure 4.3 Inter-relationship of self-management concepts 36

Figure 5.1 Execution of intervention-selection and opportune-moment-identification models in sync with the time frames associated to action plan activities . . 38

Figure 5.2 Analogy between RL and JITAI delivery mechanism 40

Figure 5.3 Agent-environment interactions during the intervention delivery process 47

Figure 5.4 Modifications on the standard eligibility traces 49

Figure 5.5 The algorithm for updating CP after each episode 51

Figure 5.6 Overall learning algorithm 61

Figure 6.1 Data processing architecture of Communication Engine 64

Figure 6.2 Internal architecture of Communication Engine 65

Figure 6.3 Internal architecture of Communication Engine 67

Figure 7.1 Episode vs intervention count vs habit strength plot 81

Figure 7.2 Person vs intervention type ratio plot 83

Figure 7.3 Difference between JITAI delivery and behavior performance times 85

Figure 7.4 Rewards per episode in the intervention selection model 86

Figure 7.5 Ratio of engaged interventions 87

Figure 7.6 Rewards collected per episode in the opportune-moment-identification model 88

Figure 7.7 Ratio of actions per action selection mechanism 90

Figure 7.8 Rewards obtained with the modified the sent-reacted variable in Eq. 5.1 94

Figure 7.9 Rewards obtained with varying scales of temporal rewards 95

LIST OF ABBREVIATIONS

aFPG	Adjusted fasting plasma glucose
A	Action set
A^{IS}	Action set for the intervention-selection model
A^{OMI}	Action set for the opportune-moment-identification model
ADA	American Diabetes Association
ADC	Accessibility decay constant
AGC_R	Accessibility gain constant due to the reminders
AGC_PB	Accessibility gain constant due to performing the behavior
AP	The set of activities included in an action plan
ATBF	Weight of behavior frequency in calculation of the accessibility threshold
ATC	Accessibility threshold constant
ATH	Weight of habit strength in calculation of the accessibility threshold
BCT	Behavior change technique
BG	Blood glucose
CP	Common policy
DP	The set of decision points
DP_{AP}	The set of decision points driven by the action plan
DP_{RW}	The set of decision points driven by the momentary context changes of the user in real world
DTH	Distraction weight due to habits
EC_a	Eligibile set of actions for a specific activity a
EC_{dp}	Eligibile set of actions in a specific decision point dp in the intervention-selection model

EMA	Ecological momentary assessment
GPS	Global positioning system
HbA1c	Hemoglobin A1c
HDC	Habit decay constant
HFTS	The set of time series data generated by the habit formation model
IS	Intervention selection
JITAI	Just-in-time adaptive interventions
M^{IS}	Markov decision process of the intervention-selection model
M^{OMI}	Markov decision process of the opportune-moment-identification model
MDP	Markov decision process
OMI	Opportune moment identification
P	Transition probability matrix
P^{IS}	Transition probability matrix of the intervention-selection model
P^{OMI}	Transition probability matrix of the opportune-moment-identification model
POWER2DM	Predictive model-based decision support for diabetes patient empowerment
QL	Q-Learning
R	Reward function
R^{IS}	Reward function of the intervention-selection model
R^{OMI}	Reward function of the opportune-moment-identification model
RL	Reinforcement learning
S	State set
S^{IS}	State set of the intervention-selection model
S^{OMI}	State set of the opportune-moment-identification model
SC	State classifier

SET	Selective eligibility traces
SMS	Short message service
TD	Temporal difference
TL	Transfer learning
TV	Tailoring variable
WCI_REM	Weight of commitment intensity in accessibility gain calculation in relation to the reminders

CHAPTER 1

INTRODUCTION

1.1 Chronic Disease Care Process Leading to Personalized Self-Management

One in four of 117 million adults in the US, had two or more chronic health diseases, also known as non-communicable diseases, as of 2012[1]. Such diseases kill almost 40 million people each year corresponding to 70% of deaths globally[2]. Projections show that the prevalence will further increase in the upcoming decades[3]. In proportion with this prevalence, chronic diseases have become a great burden on healthcare providers, especially on the primary care professionals considering also the shortage of specialists[4]. The burden is especially high on primary care professionals, who provide the majority of the services for chronic disease care[5]. According to Iyengar et al., healthcare providers have a clinic visit time of 19.3 minutes in average, leaving little time for them to analyze reams of data and review patients' various needs[6].

Fortunately, patients can strive for their health by themselves. In this respect, behavioral lifestyle patterns are important predictors of health outcomes such that patients can reduce the risk of chronic diseases by adopting healthier lifestyles. The evidence is overwhelming that physical activity and diet can reduce the risk of developing numerous chronic diseases and in many cases even reverse the existing disease[7]. Thus, during the clinic visits, despite the limited available time, patients and doctors collaboratively plan the activities that would help gaining healthier lifestyle habits on the long term[8].

Despite the action plans draw a general frame for tasks to perform, daily care is in people's own hands. Complexity of management of chronic diseases require multiple daily self-care decisions indicating that being adherent to a predetermined action plan is usually not sufficient throughout people's lives[9]. Funnel et al. further state that although the action plan is well aligned with persons' healthcare, it must also be tailored to fit their priorities, resources and lifestyle while taking multiple physiological and personal psychosocial factors into account.

1.2 Personalized Self-Management Support

Adaptive interventions have emerged in order to deal with persons' varying responses to treatments in terms of adherence as well as their evolving behavioral adoption and health outcome during the waxing and waning course of the treatment[10]. The problem with fixed interventions is that the varying intervention needs of individuals may not be met optimally by using a single, uniform composition, dosage, frequency or content of the intervention. Conversely, adaptive interventions alter these parameters across individuals, and/or within individuals across time. All these parameters vary in accordance with the intervention needs of individuals.

Employing adaptive interventions respecting to individual needs and preferences are proven to yield better outcomes in terms of quality of life and psychosocial and emotional outcomes. Existing health behavior change theories, e.g. Self Determination Theory[11], provide the scientific knowledge and background for the design of adaptive interventions. Such theories, first, define a set of conceptual input elements as the determinants of the behavior; then, examine interrelationships between these elements; and finally, predict the outputs as behaviors or health outcomes. As these models run on personal data, they output individualized outcomes. However, recent systematic review studies point out to the lack of theoretical foundation of contemporary self-management apps[12, 13]. This indicates that apps should base their self-management support strategy on literally proven theories. The lack of theoretical foundation can be explained to some extent with the gap between the conceptual basis and practical application of behavior change theories. Although they describe the behavior change conceptually, they do not specify practical methods about how to realize the change[14]. Approaches like Intervention Mapping[15] try to close this gap by introducing behavior change technique (BCT) taxonomies as a collection of practical applications of theory-based concepts.

Thanks to the technological advances it is now possible to deliver adaptive interventions wherever and whenever they are needed via mobile devices, making people accessible all the time. Interventions delivered in such a spontaneous way are called just-in-time adaptive interventions (JITAI)[16]. This concept is mainly used for interventions as a special case of adaptive interventions addressing unique and changing needs of people[17]. Mobile phone / sensing technologies enabled collection of personal data continuously. So, the JITAI concept holds enormous potential for adapting mobile phone delivered interventions to the dynamics of a person's emotional, social, physical and contextual state, so as to prevent negative health outcomes and promote the adoption and maintenance of healthy behaviors. The more frequent the data is sensed, the better capturing of dynamically changing needs, which enables provision of the type/amount of support in the right conditions[16].

1.3 A Framework for Design and Personalization of JITAIs

The main aim in this study is to deliver a reusable and expandable JITAI design and personalization framework that can be instantiated by different self-management support systems targeting different diseases/problems with different sets of intervention, triggering conditions and data sources.

Considering the design module, BCT and JITAI concepts are harmonized to provide theoretically valid self-management support to people via personalized interventions. The proposed framework provides an environment to design interventions. For the sake of clarity, it worths differentiating the user interface design of JITAIs from the design activities towards composing a JITAI by configuring their conceptual components including *decision points, intervention options, tailoring variables* and *decision rules*[18]. In this study, the focus is on the latter aspect. Offering constructs matching with the JITAI components, the proposed design mechanism facilitates JITAI design activities, as described in [4]. The mechanism is highly customizable and expandable with add-on constructs, enabling designers to customize the core capabilities to develop JITAIs tailored to a particular health problem/population. The design constructs are bound to accompanying software modules, which can also be customized and expanded, for processing of heterogeneous personal data as desired. Complementarily, a well-defined *rule definition language* is proposed to enable intervention designers defining decision rules that utilize these data integration and processing constructs.

The personalization module targets care receivers by monitoring their self-management and health-related behaviors and delivering mobile phone notifications in response to their continuously changing context during daily activities. Utilizing decision rules makes the proposed approach a rule-based system assessing the intervention triggering conditions by continuously monitoring data streams for a person including the health and behavioral observations as well as contextual information. However, purely rule-based, static systems cannot deduce opportune moments to send an intervention when the interruptibility of the person is high[19]. In addition to the momentary ones, some parameters might evolve throughout a long period of time. For example, one's preferences and perceptions about the interventions might change over time. He/she might get used to the interventions and form a kind of habituation towards similar interventions[20]; or might start feeling burdened as the number of interventions is too high and require too much cognitive resources[21]. Thus, such systems are limited in terms of personalization of the intervention delivery strategies as they do not adapt themselves in a systematic way to maintain engagement of people with interventions and improving adherence to a care program. This is a critical limitation considering that prolonged adherence is inversely proportional with the burden created by the interventions[22].

Beyond a rule-based system, a novel learning method that is closely linked with the aforementioned JITAI components is proposed. This learning method tailors intervention delivery strategies dynamically in terms of *intervention type, timing* and *frequency* using machine learning techniques, in compliance with people's action plans, changing physical / psychological contexts as well as their changing preferences over time. Thanks to their feedback loop-based learning mechanisms, dynamic systems are superior than the static ones in adapting intervention delivery according to personal values, conditions or patterns; thus, reducing the burden of interventions. Dynamic systems, as the proposed one, can tackle with cases where the static ones fail, as exemplified above. Therefore, for example, the proposed method might learn not to deliver a particular intervention for a particular person even if the triggering conditions are met considering that the user's past reactions were negative to it; or the learning algorithm might prefer postponing delivery of an intervention having the triggering conditions met but the person is driving.

Concerning the support enabled by the proposed approach, it should be noted that interventions delivered by the system do not aim to replace the care and support of healthcare professionals but aim to facilitate self-management following shared decision making, where care receivers and care providers agree on behavioral treatment goals and daily actions to pursue and follow-up until the next clinical visit of care receiver[23]. The term *action plan* is used to refer to these behavioral goals and daily actions in the rest of the manuscript.

1.4 Reinforcement Learning for Personalization of JITAIs

A reinforcement learning (RL)[24] based algorithm is proposed for personalization of JITAIs. Using RL for JITAI optimization is convenient for a number of reasons: Firstly, an RL agent learns through experience. Therefore, it does not require an initial data set, as other forms of learning methods to train the internal policy to select appropriate actions. However, it is still possible to bootstrap the learning process to make sensible choices of interventions by configuring the policy based on expert knowledge[25]. Furthermore, the RL framework is convenient for modelling people's varying context and designing a learning algorithm based on continuously changing context values.

Having an iterative nature, an RL agent performs an action in the environment it is in, observes state changes of the environment, processes the emitted rewards, update its learning model based on the reward and makes a transition to a next state. Similarly, considering the application of RL in the self-management support domain, the learning algorithm reacts to the changing context of people and updates its internal policy, aiming to discover a (near-) optimal personalized policy.

In principle, traditional RL already provides mechanisms to learn solutions for any task without the need of human supervision. However, in case RL agents begin learning tabula rasa, i.e. no prior knowledge from a domain expert is available, the number of samples needed to learn a nearly optimal solution is often prohibitive in real-world problems[26, 27]. So, learning a near-optimal policy is often slow or infeasible, especially if the experiment is being conducted includes real-world elements. In addition to long learning times, learning performance in a traditional RL system is usually asymptotical to a logarithmic curve, meaning that during the beginning phase of the learning process the actions taken will mostly be random when the state space is rarely explored[28]. Addressing these two challenges, the objective is to deliver an algorithm discovering a personalized policy quickly, which works relatively good even at the beginning of the learning process. To achieve this, two improvements on the standard RL algorithm i.e. Q-Learning[29] are introduced. The first improvement is related to better rewarding of past actions considering the effectiveness of the past actions. In this respect, the standard eligibility tracing mechanism[30] is modified by not rewarding the past actions that are perceived useless; in other words, actions that do not contribute performing the targeted behavior. Second, transfer learning method[27, 31] is utilized to reuse knowledge generated in a specific environment i.e. a specific person, across other environments. The aim is to discover common patterns that are valid for most of the people. For example, people are unlikely to engage with mobile phone notifications when they are physically active. So, such knowledge can be learnt once for a person and can be reused for others.

As will be described in detail in Sec. 5, selecting the intervention to be delivered and identifying the opportune moment to deliver the intervention are approached separately. In this respect, two dedicated RL models are employed to learn personalized patterns over these two concepts, namely *intervention-selection* and *opportune-moment-identification* models. Running in a synchronized manner, these models capture distinct components of JITAIs such that while the intervention-selection model optimizes the intervention delivery in terms of *adaptivity (i.e. time and frequency)* and the opportune-moment-identification model focuses on the *just-in-timeness (i.e. timing)* aspects.

1.5 Outline

In the rest of the manuscript, related works are presented including studies providing personalized support for self-management / behavior change and computational approaches aiming at personalization of mobile phone notifications in Sec. 2. Related works also include studies proposing alternative ways of rewarding past actions and knowledge transfer in RL environments. Background and preliminaries follow the related work. Varying concepts and techniques that are of relevance to the proposed approach are presented. The first two concepts, i.e. action plan and JITAIs, are from the self-management domain. Following them, an overview is given about RL along with the eligibility traces and transfer learning methods, which are the two concepts related to the core learning mechanism. As the last item in Sec. 3, the habit formation model that have been adopted to simulate habit-related parameters is presented. Then, in Sec. 4, the JITAI design framework is presented. Spefically, the conceptual foundations of self-management, template-based JITAI design mechanism and the rule definition language are elaborated consequtively. Sec. 5 presents the overall learning algorithm as well as the Markov Decision Processes (MDP) for intervention-selection and opportune-moment-identification models. Presentation of *Communication Engine*, follows the algorithm description. Communication Engine is the software implementing the overall conceptual approach described throughout the manuscript. Then comes the validation, which is composed of two main parts. First, the existing resources in the literature are identified that include specifications pertinent to JITAI design. Then, the way the expandable JITAI design mechanism addresses the requirements of the existing JITAI design specifications is presented. In the second part of the validation, the simulated case study is elaborated by describing simulation parameters, their differentiations among personas and hypotheses driven by these differentiations. Following the simulation design, results obtained from the simulated experiments are presented and checked against the hypotheses. The dissertation is concluded after discussing the limitations and potential improvements on the proposed approach.

CHAPTER 2

RELATED WORK

As this dissertation is intended to act as a bridge between the computer science and health behavior change fields, related works are taken into account from both perspectives. In Sec. 2.1, studies providing personalized support (in form of mobile phone-based interventions) to people suffering from chronic diseases or health problems are presented. Considering these studies, the specific interest is on their system design and computational methods facilitating the personalization of the support. By broadening the scope and getting closer to the computer science field, in Sec. 2.2, relatively more generic approaches, dealing with personalization of mobile phone notifications, are presented. Afterwards, studies related to the two main contributions of this study (customized eligibility traces and transfer learning) are presented. Specifically, studies proposing alternative methods of rewarding past actions and applying transfer learning in the RL domain are presented. Finally, in Sec. 2.5, a brief introduction is done about the related works considering the system design.

2.1 Personalized Support for Self-Management / Behavior Change

There are a myriad of recent studies providing personalized support to patients suffering from chronic diseases in their self-management activities [32, 33] or to people with health-related problems like obesity[34] or people with unhealthy behaviors like alcohol / smoking addiction[35], sedentary behavior[36] in breaking unhealthy behaviors / habits. Systematic reviews of such studies usually focus on a specific disease / problem and analyze the self-management approaches in multiple dimensions such as clinical outcomes, supported intervention types, communication mode with users and so on[12, 37, 38]. Specific interest is on studies claiming to be providing individualized and automated feedback to users since this kind of communication mode has a similar rationale with JITAIs. So, studies are not focused if they include a manual decision-making step such as a care provider evaluates the recently gathered personal data and initializes a personalized intervention after the evaluation[39, 40].

In general, the main methodology followed by studies providing automated personalized interventions is to have a set of rules to be executed upon an update on the relevant inbound data sources. Some studies use clinical guidelines as a source of such rules. Although the target audience of clinical guidelines is care practitioners, they can still be used to extract self-management rules targeting patients. For example, American Diabetes Association (ADA) recommends lifestyle recommendations including 150 minutes of physical activity per week similar in intensity to brisk walking; or reducing sedentary time by breaking after at most 90 minutes spent sitting[41]. Such conditions can be used as a goal, against which users' performance are evaluated and in turn a dynamic intervention is provided based on the evaluation. DialBetics[42] is such a system using Japan Diabetic guidelines[43] to give advice on lifestyle modifications, matched to the patient's input about food and exercise. DialBetics gather measurements twice a day. The intervention is determined and delivered right after the data gathering. Similarly, the system proposed by Liu et al. collect weekly questionnaire-based data. After assessment of the data according to Global Initiative for Asthma guidelines[44], patients receive a self-management advice immediately[45].

A similar strategy is also followed by studies that do not claim any compliance with evidence-based guidelines. The app developed by Ben-Zeev et al. prompted a questionnaire 3 times in a day within pre-determined time periods for assessment of mental status of schizophrenia patients[32]. In response to the participant entries, the system delivered a tailored intervention. Gustafson et al. introduce a set of self-management modules, one of which to track global positioning system (GPS) data and warn people suffering from alcohol addiction when they come close to a previously identified high-risk location[35]. Fioravanti et al. keep an up-to-date patient state composed of parameters like self-check inputs, drug intake, physical activity and provides automatic messaging to diabetes patients by analyzing the recent data[33].

Analysis is done via a set of pre-defined rules targeting a set of diabetic patient-related profiles such as lack of motivation for people with negative perception of the disease; comorbidities for diabetics having one or more additional disorders.

From the computer science point of view, the studies referred above can be seen as rule-based systems. Such systems evaluate the criteria to deliver an intervention always using the same rule set established addressing a specific problem once at the beginning of the care program. Most of the studies reviewed fall into this category and they neither introduce a mechanism adapting itself automatically on parameters that might change over time nor fine-tune the timing / frequency of interventions during the day. Related to the static nature of existing studies, the lack of computational models capturing complex relationships between elements of personalized behavior change in different time scales is already pointed out in [46].

2.2 Computational Approaches Aiming at Personalization of Mobile Phone Notifications

Although they are much less compared to rule-based approaches, there are also studies researching on computational approaches dealing with *adaptivity* and *just-in-timeness*, which are the two optimization dimensions targeted also in this work. Thus, both aspects are individually considered.

The adaptivity aspect
Computational approaches dealing with the adaptivity of interventions by recognizing longer term changes on individuals range from control systems engineering based dynamic modelling[47] to agent-based modeling[48] to machine learning based approaches including Bayesian network-based classification[49]. A group of studies introduce tailor-made models targeted at specific diseases / problems. For example, Chih et al. introduce an agent-based model for uncovering the predictors of food choice and obesity in the presence of cue-reward conditioning[49]. Hammond et al. use Bayesian modeling for predicting lapse status based on recovery progress and lapse history parameters[48]. Mohan et al. introduce a model-based approach maintaining a parameterized model of an individual's aerobic capability that predicts per-

formance to be able to guide users with adaptive goal setting interventions[50]. Similarly, Goldstein et al. use supervised machine learning techniques to predict dietary lapses based on contextual tailoring variables and deliver personalized interventions according to the strong predictors of the lapse[51].

Some dynamic system models, on the other hand, lay out a generic architecture capturing the models of the health behavior theories composed of interrelated variables with a mathematical representation. The mathematical representation facilitates monitoring the changes on the dynamic model variables over time and enables adapting JITAIs based on the changes[52]. While Navarro-Barrientos et al. propose a dynamical system model for Theory of Planned Behavior[53] to be used to generate weight loss related interventions[47], Martin et al. propose a model for Social Cognitive Theory[54] to perform simulations of the model using physical activity of individuals over time as a referential behavior where habituation is treated as a feature determining the behavioral response[55]. Spruijt-Metz et al. introduce a timescale separate model building on Social Cognitive Theory by establishing causal interrelationships among the variables ranging from minutely to yearly timescales[46]. This model, though, is a conceptual representation without any mathematical representation.

The just-in-timeness aspect
Mobile / sensing technologies and machine learning techniques are proposed to be used collaboratively to optimize intervention delivery concerning the just-in-timeness aspect of interventions. Towards this aim, studies try to identify interruptible moments of users where interruptibility refers to likelihood of attracting one's attention. Pejovic et al. utilize and compare a set of classifiers to predict opportune moments to deliver interventions based on variables including time, step count, location and emotions[19]. Boyer et al. use Bayesian networks to predict drug cravings based on physiological parameters including skin conductance, skin temperature, motion and pulse[56]. Morrison et al. use location, movement and time data to build a classifier for predicting users' reactions to stress management interventions[57]. Interruptibility is also studied outside the healthcare domain. Pielot et al. build a XGBoost classifier[58] using features that are formed with variables from five main groups of person data[59]. The data include communication activity, person context, demographics, state of the phone and physical activity status; and the classifier predicts

whether participants will engage with suggested content that is offered by the mobile app at the time of posting a non-specific notification. Using similar approaches, numerous other studies also provide examples of classifier-based opportune moment identification[60, 61, 62, 63].

Concerning RL, which is the specific machine learning approach used for the intervention delivery optimization, studies find utilization of RL for this problem appropriate. For example, Kelly et al. argue that reinforcement learning may enable an intelligent real-time therapy optimizing the delivery of interventions by learning over the daily psychological context to be retrieved dynamically during the day[64]. Similarly, Pejovic et al. see MDPs, which are the backbone of RL-based learning models, as a natural way to model the problem of measuring efficiency of interventions in terms of the observed behavioral outcomes[25].In [65] and [66], authors utilize RL for optimization of intervention delivery in real-time by modeling it as a contextual bandit problem[67]. These studies are similar to ours in terms of representation of the real-time intervention delivery optimization problem with RL constructs and validation of the proposed optimization algorithms in a simulated setting. Similar to the presented approach, the authors model the environment state with elements representing various momentary context parameters such as location, calendar data or physical activity status. The authors present a simplified and simulated scenario about reducing smoking for heavy smokers as an instantiation of their conceptual model. The study considers a single generic intervention type and simulates decision points of JITAIs and people's smoking urge. This approach differs from ours with respect to the learning algorithm. While the current study utilizes Q-Learning[68] supported by transfer learning for cross-individual knowledge transfer[26], the referred study utilizes an actor-critic algorithm.

2.3 Rewarding Past Actions

Eligibility traces are one of the basic mechanisms for rewarding the past actions[30]. Standard eligibility tracing has two slightly different variations, namely accumulating traces and replacing traces. In accumulating traces, eligibility value of a revisited state is increased by 1, which may end up with eligibility values larger than 1. In replacing traces case, though, eligibility values of revisited states are reset to 1. This basically means consideration of frequency and recency heuristics in accumulating and replacing traces respectively. In the current study, accumulating traces are utilized as a fast learning rate is desired for frequently visited states.

Some studies are similar to ours in terms of selective, postponed updates of the past actions considering that some of the actions taken could be suboptimal. For example, Bloch et al. propose the Temporal Second Difference Traces method so that trace updates can be performed after apparently suboptimal actions have been taken[69]. Rather than keeping track of the eligibilities in the usual sense, the method keeps track of the updates being performed. Using this information, it is able to optimize the updates as more information becomes available. Other interesting rewarding mechanisms include postponing the update of a visited state by recording the agents' last-visit experiences until the revisit of the state to improve the quality of the update [70]. Kormushev et al. propose a concept, called Time Hopping, to prevent failures (i.e. taking wrong actions) in rarely explored states[28]. The approach, though, is only valid for simulation-based studies as very little can be done in setting the environment state as desired in real-life. Nevertheless, such an approach could be used to increase the learning rates of rather rarely explored states.

2.4 Transfer Learning Strategies in RL

The survey study conducted by Taylor et al. provides a comprehensive summary about ways of applying transfer learning in the RL domain [71]. The survey classifies related studies into a set of groups based on the similarities and differences between the environments among which the knowledge could be transferred. Referring to this survey, our study is classified into the group where state and action states of the environments are the same. A subset of studies included in this group transfer whole policies among tasks. The methods utilized in this subset either learn from easy missions or break down the task to easier missions. However, the JITAI personalization problem does not have a modular structure. So, approaches working on hierarchical or modular tasks would not work on this problem. Lazaric follows a similar approach with us where state transitions are saved as tuples of source state, taken action, target state and emitted reward. His method checks the similarity of information collected in the current task to transfer the knowledge or not[72]. As elaborated in the next section, our data items, i.e. the stored data at each transition, include only state and actions, where each tuple indicates the most performed action to be taken in the associated state. In the proposed approach, such mappings are maintained in a single common policy. Nevertheless, partial policy transfers or criteria-based knowledge transfer as in [73] might be appropriate for our case as well considering that source tasks might have certain regions in their policies that have relatively more similarities with the target tasks' policies.

2.5 System Design

The closest study to ours in terms of laying out an expandable architecture for design and triggering of interventions was produced by van de Ven, Pepijn et al. [74]. Their main aim is to provide context-aware or time-based triggers for delivery of ecological momentary assessments (EMA) on mobile devices. EMAs are a means of eliciting information about a person's physical and mental state in real-time in subjects' natural environment as opposed to retrospective assessment[75]. The software they introduce, called ULTEMAT, provides capabilities for definition of EMAs along with the triggering rules. The rules can be bound to data sources available to smart phones. ULTEMAT, however, does not have a component for automated optimization of its EMA delivery mechanism. Although it is quite a recent study, the authors claim that their platform is the only one providing a flexible and versatile solution for definition and triggering of EMAs on mobile phones based on user behavior and context, which is consistent with our findings from the literature survey conducted throughout this work.

Finally, in order not to divert the focus of the study, computational scalability related aspects of the overall system as well as individual components are just mentioned briefly in Sec. 8. Nevertheless, regarding the scalability aspect, to our knowledge, no other study proposes software modules performing in a distributed manner to perform computation for self-management support e.g. parallel processing of triggering rules for interventions. There are numerous studies, though, dealing with management of data produced wearable devices, body sensors or other means on cloud-based, scalable software architectures[76, 77, 78]. More comprehensive systems as a pervasive healthcare infrastructure are also available, still with scalability capabilities[79, 80].

CHAPTER 3

BACKGROUND AND PRELIMINARIES

3.1 Self-Management of Health Problems

3.1.1 Action Plans

An action plan contains a list of planned activities to be performed by the person for achieving the targeted behavior change and associated clinical outcomes [81]. An action plan partitions a day, as depicted in Fig. 3.1, into time frames. Each activity is associated with two sequential time frames during the day. The first time frame is the period through which the activity is supposed to be performed. The second time frame follows the first one.

Such a partioning creates a set of points in time where a decision is made for delivering an intervention. To generate the action plan-driven decision points, first the set of activities (*AP*), is defined as follows:

$$AP = \{a \,|\, a \ is \ any \ distinct \ activity \ included \ in \ the \ action \ plan\}$$

Then, the set of points generated for the action plan (DP_{AP}) can be represented as, with the abuse of notation,

$$DP_{AP} = \{dp_{2i-1},\, dp_{2i} \,|\, \forall_i a_i \in AP, i > 0\}$$

The set definition above implies that there would be a corresponding decision point for each timeframe associated with a planned activity.

This structure is utilized for an initial categorization of intervention types in select-

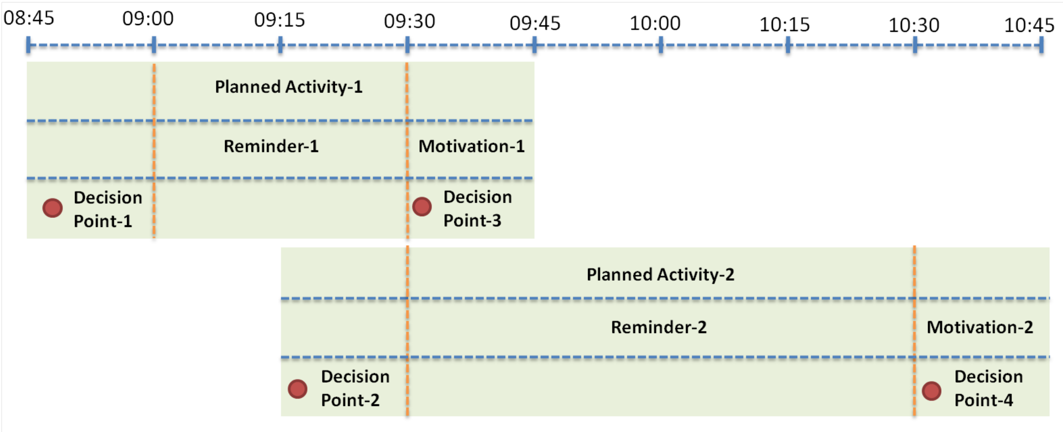

Figure 3.1: Partitioning of the daily time frame according to the action plans

ing an appropriate one for personalized self-management support. In the scope of this study, two main categories of interventions are considered namely *reminders* and *motivations*. While the reminder interventions are of use in the first timeframe, the motivation interventions are applied in the second one. Various psychological convincing techniques might be utilized while composing the intervention content. For example, a motivational intervention can benefit from social comparison and another one can use comparison of past performances of the same individual.

3.1.2 JITAI Components

Having JITAIs in the center of the proposed approach, the conceptual foundations of JITAIs are presented in this section. JITAIs, as studied in the literature, include four components[18] that are of interest for this study: *decision points, intervention options, tailoring variables* and *decision rules*

- ***Decision points*** are moments in time when a decision about the self-management support must be made. They could be periodic (e.g. daily / weekly), event-based (e.g. reaching a daily goal, or after a planned session of exercise), condition-based (e.g. if the blood glucose level is higher than a pre-defined threshold); or specified by fixed time points. The complete set of decision points (DP) is composed of DP_{AP} and DP_{RW}, i.e. $DP = DP_{AP} \bigcup DP_{RW}$. As just described DP_{AP} is driven by action plans set towards coping with the health problem. DP_{RW} contains the decision points that are driven by the changes in variables pertaining to people's momentary context. As will be described in the rest of the manuscript, DP_{AP} and DP_{RW} are utilized by intervention-selection and opportune-moment-identification models.

- ***Intervention options*** indicate possible intervention characteristics such as content, type (e.g. motivation, educational, warning), amount (e.g. dose, intensity), mode (e.g. SMS, push notification) or timing of support that can be chosen at any decision point. Intervention options have a special importance as they are the very components of JITAIs that would be optimized based on the needs of individuals by the proposed approach.

- ***Tailoring variables*** are the parameters that influence personalization of intervention options respecting individual needs. For example, location, mood or perceived habit strenght of the person can be tailoring variables to determine the best intervention options. Tailoring variables could be self-reported or passively collected data about persons' psychosocial, physiological and environmental contexts. They can also be inferred or calculated values as a result of certain data processing and analytics operations. In general, the set of all tailoring variables (TV) is defined as follows:

$$TV = \{tv \mid tv\ is\ any\ variable\ utilized\ in\ decision - making$$
$$for\ intervention\ delivery\}$$

- ***Decision rules*** are the constructs that link the former three components. Decision rules are formed with expressions conditioning on tailoring variables. They are evaluated at decision points. Upon satisfaction, each condition necessitates choosing a particular intervention option. An example decision rule could be: *Send a relaxation intervention to the user if the measured stress levels are higher than a threshold*, where change in stress level is the decision point, relaxation intervention is the intervention option and measured stress level is the tailoring variable.

3.2 Reinforcement Learning

In RL systems, learning agents observe the sequential state changes of the environment that they are in and execute actions according to the perceived state. These systems are characterized as Markovian as the knowledge accumulated during learning process retained in the recently perceived state of the environment. Design and execution perspectives are considered in order to lay out the basics for RL systems.

From the design point of view, RL systems are implemented as Markov Decision Processes (MDP). An MDP M is a tuple of four parameters such that: $M = <S, A, P, R>$, where S is the finite set of states, A is the finite set of actions, P is the state transition probability matrix including the probabilities in the form of $P(s, a, s') = P[s_{t+1} = s' \mid s_t = s, a_t = a]$, where $s,s' \in S$, $a \in A$ and t denotes the step index on the MDP. Lastly, R: $R(s,a) \rightarrow \mathbb{R}$ is the reward function mapping a state-action pair to a number. The reward indicates the desirability of the action a taken in state s concerning the maximization of the aggregated rewards in the long-run [24].

From the execution point of view, an RL system maintains a value function, Q: $Q(s,a) \rightarrow \mathbb{R}$, keeping the expected, long-term reward for the state-action pairs. At each learning step, the agent takes an action following a policy, Π: $\Pi(S) \rightarrow A$, which is a mapping from the state set to the action set. In this study, a greedy policy is used such that Π_G: $\Pi(s) = argmax_a\ Q(s,a)$, where $a \in A$, $s \in S$, meaning that at each state, the action leading to maximum long-term reward is selected. As the base RL algorithm connecting the described pieces, Q-Learning [29] is used. At each step of the learning process, Q-Learning updates the q-value of the current state-action pair. Upon taking the action a_t and observing state s_{t+1} and reward r_{t+1} at step t, Q-Learning updates q-value of the current state-action pair, $Q(s_t, a_t)$, by amount of the temporal difference (TD) error, represented with δ_t. TD-error basically represents the difference between the estimated reward at any given state and the actual reward received. Eq. 3.1 and Eq. 3.2 show the Q-Learning update procedure.

$$\delta_t \leftarrow r_{t+1} + \gamma\ max_{a \in A}\ Q(s_{t+1}, a) - Q(s_t, a_t) \tag{3.1}$$

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\delta_t \tag{3.2}$$

In these equations, α is the learning rate and γ is the discount factor which weights immediate rewards relative to future rewards.

Our main motivation for using Q-Learning is that as an offline algorithm, it allows integration of knowledge from other policies. This feature is aligned with our intention towards employing transfer learning between environments to reuse knowledge across environments. However, theoretically, the base Q-Learning algorithm requires infinitely many steps for convergence as proven in [68]. In the subsequent two sections, the preliminaries of the eligibility traces and transfer learning approaches are presented. These two methods are utilized for achieving a near-optimal policy faster than the base Q-Learning algorithm.

3.2.1 Eligibility Traces

The base Q-Learning algorithm is categorized as one-step backup algorithm as it updates the q-values of the current state-action pair by taking only the q-values of the next step into account. Algorithms using eligibility traces propagate information faster and, hence, they usually converge much faster to the optimal policies than their one-step counterparts [73, 82]. Instead of one-step backup, eligibility traces enable n-step backups by keeping the track of the trajectory of the states that the agent visited, and actions taken in the visited states. Each trace has also an eligibility value that is a weight indicating the temporal proximity of the trace to the current state. Once a reward is emitted by the environment, in addition to the current state-action pair, all the previous pairs referred by the traces are credited with positive reward or blamed with negative reward in proportion to their eligibilities [30]. At each step, all the q-values and eligibilities associated with the traces are updated according to Eq. 3.3 and Eq. 3.4, respectively.

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \delta_t E(s_t, a_t) \tag{3.3}$$

$$E(s_t, a_t) \leftarrow \lambda\gamma E(s_t, a_t) \tag{3.4}$$

In these equations, δ_t is the TD-error, λ is the trace decay parameter and γ is the discount factor.

3.3 Transfer Learning

Transfer learning is a method that allows a learning agent to leverage the past knowledge from one or more sources environments in the agent's own environment [27]. Surveys analyzing studies that utilize transfer learning in RL systems introduce a set of analysis and design activities that can be followed to apply the transfer learning properly considering the problem of interest. Firstly, the similarities and differences between the environments are analyzed with respect to state variables, action sets, goal states and transition probabilities. The next step is to choose a method for determining the source environments of which knowledge to be transferred. Simply, all environments or a specific one can be chosen; or the selection can be done according to a custom mechanism. Lastly, the granularity and structure of the knowledge to be transferred are determined. The transferred knowledge might be, for example, experience instances, policies, partial policies or q-values.

The literature also introduces a set of common evaluation metrics, which are also utilized in this study, for evaluating the outcomes of the knowledge transfer. These metrics include *jump-start, asymptotic performance, total reward, transfer ratio* and *time to threshold* [71] and they are also used while analyzing the results of the simulated case study.

- ***Jump-start*** represents the improved performance of a learning agent at the beginning of learning process compared to the base algorithm without using transferred knowledge.
- ***Asymptotic performance*** is the maximum performance that is achieved by the algorithms.
- ***Total reward*** is the amount of reward collected throughout the learning process.
- ***Transfer ratio*** is the total amount of rewards accumulated thanks to the transferred knowledge to the total amount of rewards gathered with the base algorithm without any transferred knowledge.
- ***Time to threshold*** indicates the number of steps to reach a pre-defined performance level.

3.4 Social-Psychological Model of Prospective Memory and Habit Development

As a person performs a particular behavior more and more, s/he forms habit for that behavior enabling the person to spend less and less cognitive resources to remember and perform the behavior [83]. Tobias introduces a social-psychological model of prospective memory and habit development, which was also tested with empirical data[84]. The model is related to this study in terms of simulation of the personal real-life conditions. Specifically, for the sake of having a theoretically valid simulation, the model is benefited for simulation of habit-related behavior parameters that are in turn related to performance of the targeted behavior.

The model introduces a set of principles explaining the effect of memory aids on the behavior performance. In this respect, the model introduces *accessibility* and *accessibility threshold* as the main determinant concepts for remembering a behavior. The accessibility concept is described as an indicator of ease of remembering the behavior. The model states intensity of commitment to the behavior, forgetting, behavior performance, events and situational cues as the factors influencing the accessibility. For example, while forgetting decreases the accessibility; performing the behavior and availability of situational cues increase the accessibility over time. Accessibility threshold is the value that the accessibility must reach for a memory content to be remembered. According to the model, the threshold value depends mainly on the available cognitive resources, behavior performance frequency and habit strength parameters. Among these parameters, habit strength has its own dynamics such that while it decays in proportion to its current value if the behavior is not performed, it increases by performing the behavior.

The model is instantiated with two external parameters namely the behavior frequency and the commitment intensity. While commitment intensity is defined as the strength of any form of internal pressure felt by a person to perform a behavior, behavior frequency is defined as the ratio of number of times executing the behavior to the total number of opportunities to perform a behavior, within a certain period in the past. According to the model, the time to form a habit for the targeted behavior is inversely proportional with these parameters. That is, the higher the commitment intensity and/behavior frequency, the shorter time needed to form habit or vice versa. These parameters are configured while simulating habit formation processes of the hypothetical personas in the scope of the validation activities.

The dynamics described above are represented with an iterative mathematical model. At each iteration, the values of the *accessibility, habit strength, behavior frequency* and *salience of situational cues* (which is represented numerically) are updated. The habit formation model is iterated in sync with action plans. Time-series data set generated by the habit formation model can defined as:

$$HFTS = \{<accessibility, habitstrength, behavior frequency, \\ salience_1, ..., salience_n >_i | \forall_i dp_i \in DP_{AP}\}$$

In the set definition above, n corresponds the number of intervention types that are designed by the behavior change experts in a specific self-management support program.

Upon comparing the current values of accessibility and accessibility threshold, the model provides its output as an indicator of whether the person would remember performing the behavior or not. The assumption is that if the behavior is remembered, it would also be performed by the person; otherwise, it would not.

CHAPTER 4

JITAI DESIGN

In this chapter, first, the conceptual foundations of self-management are presented by explaining the relationships between several concepts, such as behavioral goals, action plans or interventions, pertaining to the self-management of health problems. The main motivation is to show how JITAIs are related to these self-management concepts. Following the conceptual foundations of self-management, the template-based JITAI design mechanism is described by mapping its elements to the JITAI components that are studied in the literature. Lastly, the *Rule Definition Language* is presented by elaborating its built-in constructs. The design module of the proposed framework has also been published in [85].

4.1 Conceptual Foundations of Self-Management

Self-management conditions change during a care process. New goals might be set, action plans might be updated or new personal values might be considered. To be able to have a generic support system for self-management, meeting the requirements of such changes, either appropriate interventions should already be in place or the system should be easily expandable to introduce new interventions. So, this implies that JITAIs would be adapted upon such changes to deliver the appropriate support. Such a goal requires a structured representation of the relationship between JITAIs and other self-management concepts that are subject to change. As will be presented below, this relationship, in turn, shapes the intervention delivery mechanism in enumerating the time points for initiating an intervention and filtering a set of eligible interventions at those points among the complete set of interventions. Fig. 4.1 shows

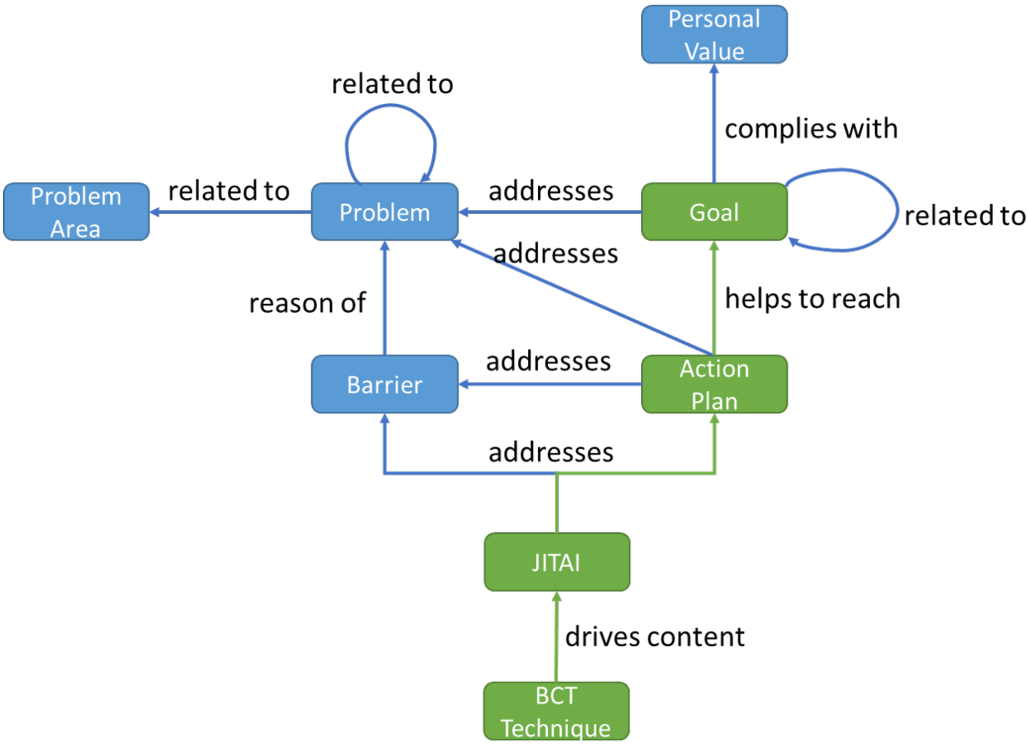

Figure 4.1: Inter-relationship of self-management concepts

the relationships of self-management concepts.

The need for self-management primarily arises from a chronic disease (e.g.diabetes) or a medical condition (e.g. obesity). Self-management activities are performed to deal with a specific *problem* like too few blood glucose measurements related to the main disease or condition. Problems are related to a *problem area*, which is a health-related behavior such as, but not limited to, blood glucose monitoring. carbohydrate intake, physical activity or stress. (Examples are given in relation to the diabetes diseases as the first potential real-life case study will be in this area). Problems might be related to other problems that must also be addressed e.g. insufficient knowledge about the usage of medical devices. *Personal values* describe what is important for people, serving as a compass guiding them throughout their lives. *Barriers* are obstacles for many people in meeting realistic goals and activities for self-management and hence, in attempting to move towards their values [86]. They are usually the main causes of problems e.g. forgetting or feeling frustrated over lack of success. The concepts so far, i.e. the ones represented with blue boxes, were the concepts that reflect the present status of people pertinent to the main disease/health condition.

On the other hand, the concepts represented with green boxes can be seen as counter-measures against the major health problem. In general, the counter-measure concepts are initialized in a way that meet the requirements of problem solving strategies and also respect the personal values and conditions of people. The self-management support aimed via the delivery of digital interventions does not aim to replace the care and support of healthcare professionals but aim to facilitate self-management after shared decision making, where care receivers and care providers agree on behavioral treatment goals and daily actions to pursue and follow-up until the next clinical visit [23]. In the light of this approach, goals specify the targets to be achieved as an indication of resolving the problems. Goals can have different timescales e.g. weekly, monthly and they can be related to each other. For example, a short term behavioral goal (e.g. minimum 150 minutes of daily physical activity) can be set to reach a longer-term health status goal (e.g. %0.7 decrease on HbA1c in 6 months). Action plans are associated on one hand with barriers and problems and on the other hand with goals, where the goals are resolving indicators of corresponding problems. Action plans contain a list of planned activities to be performed by the person for achieving the targeted behavior change and associated outcomes [81]. Specifically, each activity is associated with a particular problem meaning that performing the activity helps overcoming that problem. Having their own objectives, activities may require dedicated methods to support people, which would be realized with appropriate interventions. JITAIs, in this respect, are intended to be in place whenever needed with the required characteristics. Lastly, BCTs recommend varying psychological methods to overcome a problem. Therefore, the content of a JITAI is defined in conformance with the recommendations of BCTs. Behavior change taxonomies [87, 88] are sources of behavior change techniques (BCT) that can be used to conceive theory-driven interventions

4.2 Template-Based JITAI Design

The template-based design mechanism enables configuration of interventions targeting specific health problems in compliance with the JITAI framework presented in [16]. The mechanism facilitates intervention design for behavior change scientists and intervention designers by enabling configuration of four JITAI components namely; *decision points, intervention options, tailoring variables* and *decision rules*. While the template contains elements that match with these JITAI components, it has also additional elements setting up the relationship between JITAIs and action plans as laid out in the previous section. Below, the elements supported by the JITAI-design template are described.

The *description* element is for convenience of intervention designers, letting them to state the objective and reasoning of the JITAI along with the conditions suitable for delivering the intervention in a human-readable manner. The *targeted behavior* element links the JITAI to a problem area e.g. *blood glucose monitoring* and therefore to the problems associated to the problem area. As presented in detail in Sec. 5, this relationship is utilized for narrowing down the complete list of interventions as required by the learning algorithm towards adaptation of intervention types.

The *decision point* element provides the capability to define *event-based* and *time-bound* points when a decision would be made to initiate an intervention (e.g. deliver a notification via mobile phone) or not. Time-bound points could be specified as specific points (e.g. at 8:00AM and 9:00PM) or in a periodic way, optionally during a certain time-frame (e.g. at each 30 minutes between 10:00AM and 6:00PM). Event-based points capture user-initiated points, which could be manual (e.g. when the person asks for care provider support) or bound to a change in a certain tailoring variable (e.g. when the daily step count exceeds 10000).

Action plans are another source of event-based points in the following manner: Each planned activity in the action plan is supposed to be performed within a time frame(e.g. first blood glucose monitoring activity should be performed between 9:00AM and 11:00AM in the morning). In this respect, the proposed framework generates event-based decision points related to the performance of a specific planned activity. As might be recalled, an example of decision points generated for two planned activities in an action plan was depicted Fig. 3.1. Two decision points are created for each activity such that the first one is set for an upcoming activity before the activity time frame. This point might fall also into the activity time frame if the behavior has not been performed yet. The second point is set for after the activity time frame.

These two types of decision points are called *upcoming_action*and *post_action*. They restrict the set of whole set of interventions keeping only the eligible interventions according to their categories, which are defined in the next paragraph.

The *category* element is closely linked with the decision points. It serves for two purposes. Besides indicating the type of interventions, it is also an indicator for interventions' decision points such that only interventions categorized as *reminder* would be considered in the upcoming_action decision points and only *motivation* interventions are considered in the post_action decision points.

A mapping (M) is established between each activity a defined in an action plan $(a \in AP)$ and an eligible set of interventions (EA_a) such that $M : A \rightarrow EA_a, a \in AP, EA_a \subset A^{IS}$. Therefore, at each point, an action is_a_t is selected such that $is\_a_t \in EA_a$. This mapping is established via the *targeted behavior* component of the JITAI design template. Each JITAI instance is also associated with a behavior. On the other hand, each a decision point is also related to an action plan activity, which is prescribed for a specific behavior. Therefore, a relation between the activities decision points and eligible interventions is established via the common behavior addressed by these two concepts. This relationship is illustrated in Fig. 4.2.

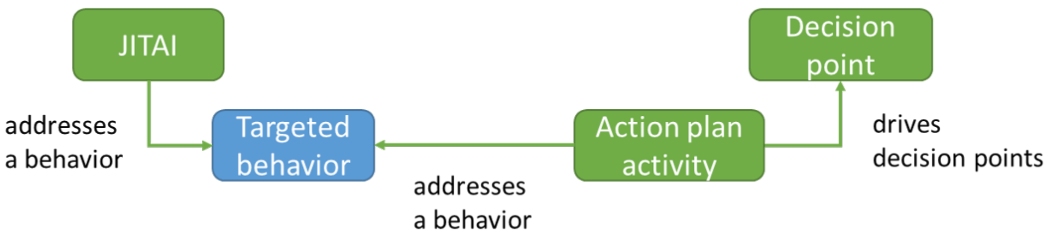

Figure 4.2: Relationship between interventions and decision points

Furthermore, the EA_a set is filtered according to the type of the decision points. That is, if a decision point prior to the activity, only the *reminder* interventions are selected. Otherwise only the *motivation* interventions become potential candidates to be selected. Let's define the final eligible set of actions at a decision point as EA_{dp}, which will be referred in the next chapter.

The *behavior change technique* and *content* enable specification of the intervention options. These two elements are related to each other such that the designers are supposed to determine an appropriate content for the intervention as suggested by the behavior change technique. In this way, several interventions, implementing different psychological methods, can be defined for the same purpose. For instance, a motivational intervention can be instantiated by providing information about the activities to be performed, comparison-by-self considering the person's past performance or comparison-with-others [88].

The *rule* element corresponds to the *decision rule* component of JITAIs. It incorporates the *tailoring variables* as operands in the decision rules. Tailoring variables can be any raw / processed data or inferred knowledge obtained from people's contexts related to their environment, their physical or medical conditions or devices they use. The rule construct is elaborated in the next section by presenting its role in the *Rule Definition Language*.

Last but not least, Nahum-Shani et al. introduce two additional concepts namely, *distal outcome* and *proximal outcome* to represent goals to be achieved by care receivers with the support of JITAIs[18]. Distal outcomes usually represent the ultimate goals as primary clinical outcomes like losing weight or having lower levels of HbA1c. Proximal outcomes intend to represent relatively short-term goals through which the effectiveness of interventions can be measured. These two concepts are not the direct properties of a JITAI. Rather, they are captured by the goals defined in action plans. Each intervention instantiated with the proposed template is associated with actiona plan goals via the *associated goal* element. In this way, an intervention can be linked with one or more goals, therefore, with proximal and distal outcomes. As described in the Sec. 4.3, a set of built-in constructs are provided for measuring the effectiveness of interventions considering the targeted outcomes.

Table 4.1: Elements of JITAI design template

JITAI Component	**Template Element**	**Element Example**
-	*Description*	One achieves a daily, weekly or monthly blood glucose monitoring goal consequtively, and the system motivates her/him to maintain the behavior
-	*Targeted Behavior*	Blood glucose monitoring
Decision Points	*Category*	*Reminder* or *motivation*
Decision Points	*Decision Points*	*event_based* points driven by the planned activities in action plans, enumarated as *upcoming_action, post_action* *event_based* points conditioned on tailoring variables e.g. step count exceeds 10000 *time_bound* points that could be specific time points or time points specified via period and/frequency base points
Intervention Options	*Behavior Change Technique (BCT)*	Providing rewards contingent on successful behavior(derived from CALO-RE taxonomy [88])

Continued on next page

Table 4.1 – *Continued from previous page*

JITAI Component	Template Element	Example Instantiation
Intervention Options	*Content*	"*en*": " Well done you are doing a great job! You successively achieved your BG monitoring goal for last *${streak_value} ${streak_temporal}*s.", "*es*": " Bien hecho, está haciendo un gran trabajo! Su objetivo de monitorización de la glucosa ha sido alcanzado exitosamente durante los últimos *${streak_value} ${streak_temporal}*. ", The example has two placeholders that are *streak_value* and *streak_temporal*. While the former placeholder represents the number of sequential temporal periods in which the person reached the blood glucose monitoring goal, the latter specifies the temporal period e.g. days, weeks or months.
Decision Rules / Tailoring Variables	*Rule*	[goal.monthly=*ACHIEVED* and goal.monthly[1]=*ACHIEVED*, goal.weekly=*ACHIEVED* and goal.weekly[1]=*ACHIEVED*, goal.daily=*ACHIEVED* and goal.daily[2]=*ACHIEVED*].
Distal / Proximal Outcomes	*Associated Goal*	• Monitoring blood glucose levels three times a day • Minimum of 8000 steps per day • 7% HbA1c at the end of three months

4.3 Rule Definition Language

This work is closely related with the activities performed in the POWER2DM(Predictive Model-Based Decision Support for Diabetes Patient Empowerment) project[1]. The objective of POWER2DM is to develop and validate a personalized self-management support system for diabetes patients. A multidisciplinary discussion between computer scientists and cognitive behavioral psychologists in the project revealed that the interventions must meet the following requirements to be effective. They should 1) ensure clinical safety, capturing the specifications from clinical guidelines, 2) be in line with personal self-management goals and actions as planned in clinical visits, 3) be in line with health behavior change theories and 4) conform to people's preferences in order not to create burden with irrelevant notifications sent in inappropriate times[89]. Aiming to be sufficiently expandable towards meeting these requirements, a rule definition language is proposed with the Backus–Naur form notation [90] as depicted in Listing 4.1. The rules specified via this language are set as the *rule* element of JITAI templates.

Listing 4.1: Grammar of the rule definition language

```
<tailoring_variable>::=<tailoring_variable>|
                           <tailoring_variable> <temporal>
<temporal>        ::=<temporal> | <temporal> "["<index>"]"
<rule>            ::=<tailoring_variable> <operator> <number>
                  <tailoring_variable><operator><tailoring_variable>
<rule_list>        ::=<rule> | <rule> <boolean_op> <rule_list>
<rule_list_list>  ::=<rule_list> | <rule_list>","<rule_list_list>
<decision_rule>  ::="["<rule_list_list>"]"
```

Tailoring variables are data integrating and processing constructs, with corresponding software modules. They transform raw data aggregated from external sources to an actionable form to be actioned by the modules of the decision making and learning pipeline of JITAI personalization. Tailoring variables are bound to physiological, psychological or environmental contexts of people. They include, but not limited to, measurements obtained from medical devices, raw data sensed from wearables and

[1] http://www.power2dm.eu/

phone sensors as well as actionable knowledge inferred from factual data via various data analysis operations. Goal-related variables let gauging the effectiveness of interventions. However, each goal type has its own internal logic for such an evaluation, which requires specialized methods. Tailoring variables can be specialized by suffixing sub-variables. For example, *adherence:bgm* outputs the adherence for the blood glucose monitoring behavior given the corresponding software module is in place.

Individual variables form a *person state* which is updated continuously each time one of the variables change. The information aggregrated inside the person state is consumed by the decision rules as described below and by the learning algorithm as described in Sec. 5.

The *temporal* construct is used to evaluate a tailoring variable considering a specific time interval. It can be set to *daily, weekly* or *monthly* to get the average value during the specified interval. It can also be set to *best, weekly-worst* kind of specifiers to get peak values for the desired variable. The temporal construct can also be assigned with an index that allows data retrieval for a specific period in the past. The following rule expression can be given as an example containing the design constructs introduced so far: *stress.monthly < stress.monthly[-1].* The expression is interpreted as follows: The average stress value in the current month should be less than the average stress value during the last month.

The proposed design mechanism is expandable via a bottom-up manner. At the finest level, any type of data sources, either static or streaming, can be integrated for better sensing of people's internal and external contexts on varying time scales. A multi-dimensional data space can be defined for a person with data types including but not limited to demographic information, self-reported information via questionnaires or mobile app, measurements obtained from medical devices, contextual information sensed from wearables and phone sensors as well as actionable knowledge inferred from factual data via various data analysis operations. Inferred knowledge may include, for example, adherence to specific activities prescribed in action plans, relapse risk of unhealthy behaviors or enumerated representation of the person's geographic location. Following the data integration, it is possible to define custom data processing modules to be used throughout intervention delivery decision making. Such modules could for example calculate the average blood glucose values during the last month or compare measurements of a person with the rest of the population. The topmost level at the expandability approach is definition of new interventions by reusing the data processing modules. This is realized via the rule definition language, specficially by tailoring variables, which enable referring to the data processing modules in decision rule configurations of JITAIs. Appendix A contains the complete list of built-in constructs included in the rule definition language, as specifically used in the POWER2DM study/real word care program, for which the details are presented later on.

Having elaborated the JITAI design mechanism and the rule definition language, a more comprehensive example is presented in Fig. 4.3. The figure shows a set of complementary interventions addressing the blood glucose monitoring problem. From left to right, decision points are the first branching node for interventions. Each particular value of the decision point makes only a subset of interventions available for delivery. Then, rules including conditions on tailoring variables further restrict the eligibile intervention set further based on the specified conditions. For instance, intervention 7 and 8 are considered for delivery only if the weekly or monthly goal is not achieved. Each intervention is associated with a theory-based behavior change technique, which in turn drives the content of the intervention. Such sets of interventions could be defined based on the combinations of decision points, goals and conditions.

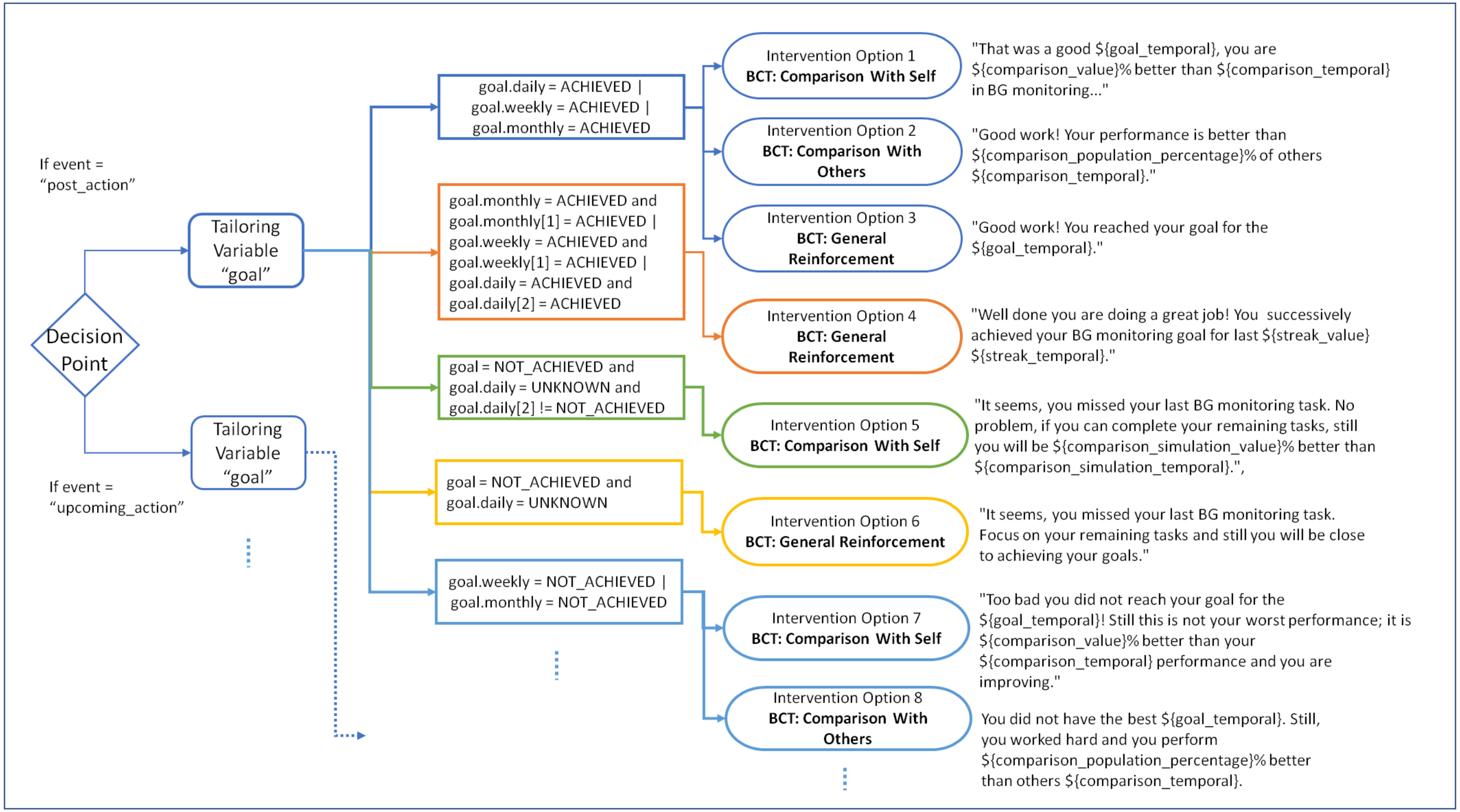

Figure 4.3: Inter-relationship of self-management concepts

CHAPTER 5

THE JITAI PERSONALIZATION ALGORITHM

In this chapter, the JITAI personalization algorithm is presented. Personalization occurs by learning person-specific patterns on intervention delivery. Reinforcement learning (RL)[24] method is used as the backbone of the proposed algorithm, learning such patterns. Considering that the aim is to learn a personalized strategy on various aspects of intervention delivery, first, an analogy is established between the feedback loop-based mechanisms of intervention delivery and RL, which is partially described in a previously published article [91]. The two concepts are further aligned by distributing the JITAI elements over two RL models. It is claimed that the RL models are able to capture the dynamics of the JITAI elements and personalize them in terms of *adaptivity (i.e. time and frequency)* and *just-in-timeness (i.e. timing)* aspects of JITAIs only if they are employed simultaneously. The first model, called the *intervention-selection* model, adapts the intervention delivery strategy considering that the person still needs to be supported via interventions or s/he does not need extrinsic support anymore e.g. the person might have formed habit on the targeted behavior. In the meantime, the model also learns personal preferences on intervention types throughout the overall learning process. The second one, i.e. the *opportune-moment-identification* model, learns the most opportune moments throughout people's daily lives to deliver interventions.

Before elaborating the conceptual and technical details of the learning models, presenting an overview of the overall algorithm would help reader to better follow the details presented in the subsequent sections. The proposed algorithm is an RL-based, iterative algorithm that aims to take the best action at each iteration. It works in sync with people's action plans that are created collaboratively by care receivers and care providers to cope with a health problem. Fig. 5.1 shows how the execution of the RL models are projected onto the time frames of action plan activities. $\forall dp, dp \in DP_{AP}$, i.e. for all decision points of intervention selection model, a state transition occurs such that the learning agent travels from is_s_1 to is_s_n. At each decision point, the intervention selection model selects an intervention for delivery. Similarly, $\forall dp, dp \in DP_{RW}$, the opportune-moment-identification model makes state transitions starting from omi_s_1 to omi_s_n. Each state transition occurs when there is a change in the momentary context of the person. However, the opportune-moment-identification is disabled in cases no particular intervention type is selected by the intervention-selection model. For example, considering Fig. 5.1, no intervention type is selected in is_s_2. Therefore, there is no state transition of the opportune-moment-identification model throughout the second time frame. At each decision point, the opportune-identification-model decides to deliver the selected intervention or not.

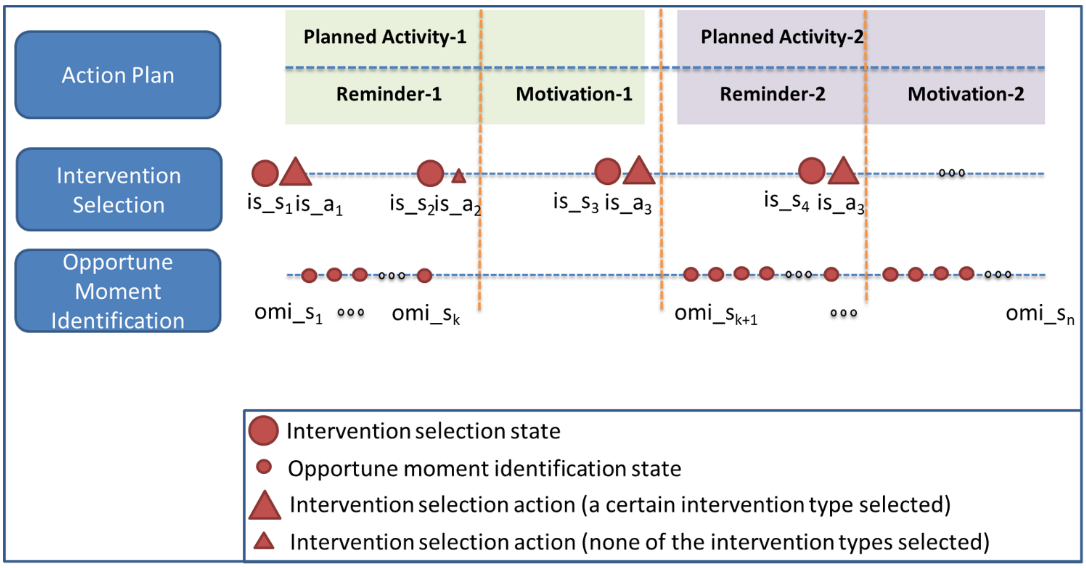

Figure 5.1: Execution of intervention-selection and opportune-moment-identification models in sync with the time frames associated to action plan activities

However, each model has its own iteration patterns, actions to take or state compositions. Furthermore, at each iteration, each model follows distinct set of procedures to update its state parameters or to select an action. Sec. 5.1 is dedicated to introduce the conceptual differences in the RL setups of the two models by setting out from an analogy between RL and JITAI framework. Deriving from this analogy, the opportune-moment-identification and intervention-selection models are formalized with two MDPs in Sec. 5.2.1 and 5.2.2 respectively. As mentioned, the mechanics of the models, therefore the MDPs, differentiate from each other. Specifically, the opportune-moment-identification model utilizes the two aforementioned optimization techniques, i.e. customized eligibility traces and transfer learning. These two concepts address the *jump-start* challenge attributed to RL algorithms. The technical details of the eligilibity traces and transfer learning techniques are presented in Sec. 5.2.1.1 and Sec. 5.2.1.2 respectively. On the other hand, the intervention-selection model embeds an evidence-based, mathematical model to update some of the internal state parameters pertaining to habit formation. As previously described, the habit formation model is also an iterative model running in sync with the intervention-selection model. Similarly, in Sec. 5.2.2, how the habit formation model is integrated with the simulation of habit related state parameters at each iteration i.e. state transition is presented. Finally, the complete sequence of the algorithm steps is presented in Sec. 5.3 connecting all the optimization methods and showing interlinked execution of the intervention-selection and opportune-moment-identification models.

5.1 Alignment of Self-Management Concepts with RL

Alignment of the static conceptual elements of JITAIs and dynamics of intervention delivery mechanism with RL is a prior, necessary step to be able to define the RL models with complete details. The objectives of establishing such an alignment include: to show why RL is a convenient approach to solve the problem of JITAI personalization; how the RL models capture the theoretical characteristics of JITAIs; and how they work in harmony on one hand with these theoretical foundations and on the other hand with action plans prescribing practical measures to tackle with health problems.

5.1.1 Analogy of Intervention Delivery and RL

JITAI personalization problem is approached from two perspectives. Selection of the correct intervention type and selection of the proper moment to deliver interventions are treated as two sub-problems to be dealt with separately, in a person-specific way. Although these concepts are captured with distinct analogies, the alignment depicted in Fig. 5.2 is applicable for both cases. Below, the common and distinctive characteristics of each analogy are discussed.

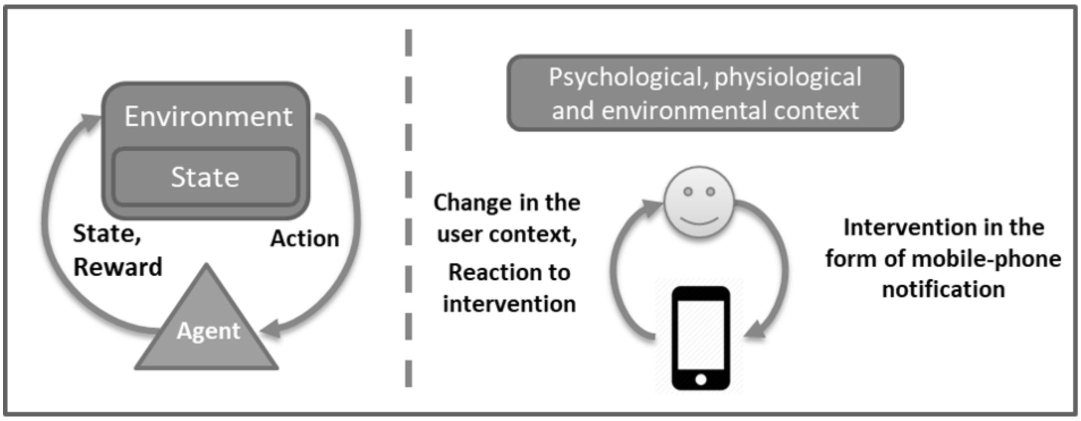

Figure 5.2: Analogy between RL and JITAI delivery mechanism

Common characteristics in both analogies: As stated previously, a traditional RL setup includes two main concepts namely the environment and agent. As depicted in Fig. 5.2, the person, as the observed entity with an associated internal state, corresponds to the environment concept of the RL mechanism. The person state is composed of parameters pertaining to the person's psychosocial, physiological and environmental contexts. Instantaneous values of these parameters represent the state that the environment is in. Changing context of the person is modeled as a series of state transitions of the environment. The mobile app, as the learning agent observing the person, takes relevant actions considering the current person state at each state change. The engagement and reaction of the person to delivered interventions as well as performance associated to the target behavior are emitted as a reward.

The two analogies differentiate in four parameters namely, *state parameters, state transitions, action sets* and *reward functions*. Considering that these are also the elements of an MDP, elaboration of the analogies reveals the details of the corresponding RL environments.

Specifics of the intervention-selection - RL analogy: The initial differentiating factor of the intervention-selection - RL analogy is the parameters composing the person state. The person state of the intervention-selection model mostly contains parameters that have long-term timescales. Because, selection of the best-matching intervention type, where selecting none of the available interventions types is also an option, can be learnt via the long-term parameters such as preferences on intervention types, past performance or habit strength than momentary parameters. The next differentiating factor is the state transitions. The intervention-selection model captures the dynamics of an action plan by evaluating the necessity of delivering an intervention for each planned activity as shown in Fig. 3.1. The model decides on the type of the intervention to be sent. Therefore, state transitions of the RL environment correspond to the decision points driven by the action plan.

At each transition, either intervention types that are applicable for the specific activity is selected as the action or no type is selected at all (meaning that no intervention would be sent) for that activity. Finally, the emitted reward has two meanings depending on the selected action. In case of a certain intervention type is selected, the reward indicates the acceptance/non-acceptance of the intervention type by the person. Otherwise, i.e. in case no intervention delivery action is taken, the reward indicates the necessity/non-necessity of an intervention for performing the behavior.

Specifics of the opportune-moment-identification - RL analogy: In this case, the state parameters are mainly composed of momentary context parameters like time, physiological status or location as the aim is to identify an opportune moment to deliver an intervention during the day. So, a state transition occurs each time one of these parameters change. The available actions at each state include delivering an intervention or not. The emitted reward signal hints about the suitability/non-suitability of the moment for the person's engagement with the intervention.

5.1.2 Distributing the JITAI Components over the Two RL Models

As introduced before, the template-based design mechanism addresses the four JITAI components as follows: *decision points, intervention options, tailoring variables* and *decision rules*. Opportune-moment-identification and intervention-selection models are also able to capture the dynamics of these elements. As a result, the learning procedure is able to optimize the intervention options, which are the adaptable components of JITAIs.

Intervention options are mainly related to the content, delivery mode or timing of JITAIs. Therefore, such parameters can be seen as characteristics of a JITAI. In the presented alignment, intervention options related to content are captured by the intervention-selection model. On the other hand, the opportune-moment-identification model deals with the timing of interventions.

Tailoring variables capture peoples' specific experiences (e.g. high levels of stress) or contexts (e.g. prolonged sedentary behavior) and moderate the decision making towards dealing with the captured values via an appropriate intervention option. Therefore, tailoring variables correspond to the state parameters that are used for representing the state the environment is in. An instantiated state represents logically conjunction of tailoring variables composing the state. For example, let's assume that the state is composed of two parameters: 1) duration of being sedentary in minutes and 2) a binary value indicating that the current stress value is higher than the weekly average. Then, a state with (30, yes) values indicates that the person is sedentary in the past 30 minutes and his/her stress level is higher-than-average. In the proposed approach, tailoring variables are distributed over the two RL models. While the tailoring variables pertaining to long-term parameters are mapped to the states of the intervention-selection model, the tailoring variables related to momentary parameters are captured by the opportune-moment-identification model.

Decision points of JITAIs correspond to state transitions of the RL environments. Therefore, the intervention-selection and opportune-moment-identification models have their own decision point patterns. Fig. 5.1 has already showed the decision point patterns for both models. The eventual RL algorithm works in sync with decision points of both models. At each decision point (i.e. at each state change), it produces a relevant reward according to the underlying model.

Decision rules link intervention options, decision points and tailoring variables. As just discussed, intervention options are captured by RL actions, tailoring variables are captured by state parameters and decision points are captured by the state transitions. And lastly, a learning agent's policy accumulates state-action mappings where each such mapping indicates a specific action to be taken in a specific state. Each distinct state can be considered as a set of decision rules based on the conditions constructed with the instantaneous values of the tailoring variables. Considering the example above, the decision rule can be interpreted as logical conjunction of the two tailoring variables. That is, when the person is sedentary for more than last 30 minutes and has a higher-than-average stress level.

Rewards collected by the learning agent during the learning process is a performance indicator for the intervention delivery strategy considering the goals set for the targeted behavior. As presented earlier, the goals are representatives of the *proximal/distal outcomes* of the JITAI framework. Therefore, an indirect relation between rewards and proximal/distal outcomes can be established such that the more rewards that the learning agent collects, the more the targeted outcomes are achieved.

The agent optimizes its policy over time by learning from people's experiences. As it visits different states, it learns how to behave in different conditions and finds a (near-) optimal strategy that suits the user best. With this base approach, though, the learning algorithm needs to consider all tailoring variables inside the person state against the complete set of actions (i.e. intervention types), which would require a long learning time. Instead, decision rules associated to individual intervention instances are used to limit the complete action set by keeping only the interventions of which rules are satisfied at each decision point.

5.2 RL Models

Driven by the analogies above, dedicated RL models are employed for personalization of the intervention-selection and opportune-moment-identification concepts. In this section, the formal definitions of these models are presented. It is assumed that people are fully observable thanks to the directly sensed or inferred contextual values required throughout the learning process. This makes using MDPs appropriate to formally describe the RL environments. An MDP M is characterized with a tuple consisting of four parameters (state set, action set, probability matrix and reward function) represented as $M = <S, A, P, R>$. Below, the intervention-selection and opportune-moment-identification models are presented in terms of these four MDP components.

5.2.1 The Opportune-Moment-Identification Model

The MDP instantiated for the opportune-moment-identification model is represented with:

$M^{OMI} = <S^{OMI}, A^{OMI}, P^{OMI}, R^{OMI}>$.

Considering that the opportune-moment-identification model aims to capture person-specific patterns over momentary data, the person state includes parameters pertaining to physiological, psychological or environmental context of the person. Each state s^{OMI} such that $s^{OMI} \in S^{OMI}$ is composed of six parameters as defined below:

s^{OMI} = <*time, location, physical_activity_status, phone_screen_status, emotional_status, number_of_interventions_sent_for_planned_activity*>

The time required to reach a (near) optimal policy with an RL solution is inversely proportional with the size of state and action sets. Therefore, state parameters with continuous values are avoided, but they are modeled as discrete parameters with a range on the natural numbers set, as described below:

- ***time*** $= \{x: x >= 0 \text{ and } x < 95, x \in \mathbb{N}\}$. This parameter is used to represent the time during the day. The time value is generalized to represent only the quarterly periods. Its value is calculated with the following equation: $time = h*4 + \lfloor m / 15 \rfloor$, where hour h is defined as $h = \{h: h >= 0 \text{ and } h < 24, h \in \mathbb{N}\}$ and minute m is defined as $m = \{m: m >= 0 \text{ and } h < 60, m \in \mathbb{N}\}$.
- ***location*** : Location represents whereabouts of the person with three enumerated values namely *Home*, *Office* and *Other*.
- ***physical_activity_status:*** This parameter is used to enumerate the physical activity of the person with *Sedentary*, *Indoor_Activity*, *Walking*, *Running* and *Driving*

alternatives.

- ***phone_screen_status:*** Enumerated value of phone screen status that can be *On* or *Off*.
- ***emotional_status:*** Enumerated value representing the emotions that a person may have including *Neutral*, *Happy*, *Excited*, *Relaxed*, *Angry* and *Stressed* alternatives.
- ***number_of_interventions_sent_for_planned_activity:*** This parameter keeps the number of delivered interventions for the active activity defined in the action plan.

The action set of the opportune-moment-identification model(A^{OMI}) includes two elements: *Deliver_Intervention* and *Deliver_Nothing* that are respectively used for delivering a certain type of intervention or not delivering any intervention. The set is formally described as follows:

A^{OMI} = *{Deliver_Intervention, Deliver_Nothing}*

R^{OMI} first considers the action taken at step t (omi_a_t). The algorithm generates a relatively small negative reward denoted with the not_sent variable in order to prevent the learning algorithm getting stuck at a local minima, which is caused by never selecting an intervention. In cases where an intervention is delivered ($omi\_a_t = Deliver\_Intervention$) the algorithm initiates the reward by considering the person's reaction to the intervention. Rewards $sent\_reacted$ and $sent\_not\_reacted$ are generated for reacting ($reacted_t = true$) and non-reacting ($reacted_t = false$) to the intervention. Initial values for $sent\_reacted$ and $sent\_not\_reacted$ are set to 1000 and -2 respectively. This reflects the higher importance given to discovery of a state that is suitable for intervention delivery, that leads to engagement with the intervention; than discovery of a state resulting without any engagement. In cases where an engagement does not happen the negative reward is multiplied with the number of successive attempts of intervention delivery. The aim is to minimize the burden of interventions on people. In cases where engagement happens, the reward is further updated according to the temporal proximity of the intervention delivery to the actual engagement time with the intervention ($difference_t$). Three ranges are defined into which the temporal difference (td) of delivery and engagement can fall. The ranges can be represented as: *0 <= td <= 30, 30 <td <= 60 and td >60.* The lower the temporal difference, the higher the reward. Initial values picked for these cases are 2, -1 and -2 respectively. As a summary, while Eq. 5.2 shows calcuation of the reward generated for the temporal proximity, Eq. 5.1 shows calcuation of the eventual reward generated at each step of the opportune-moment-identification model.

$$omi\_r_t = \begin{cases} \textit{not\_sent}, & \text{if } omi\_a_t = Deliver\_Nothing \\ \textit{sent\_not\_reacted *number\_of\_attempts}_t, & \text{if } \begin{array}{l} omi\_a_t = Deliver\_Intervention, \\ reacted_t = false \end{array} \\ \textit{sent\_reacted +temporal\_reward}_t, & \text{if } \begin{array}{l} omi\_a_t = Deliver\_Intervention, \\ reacted_t = true \end{array} \end{cases} \tag{5.1}$$

$$temporal\_reward_t = \begin{cases} \textit{long\_difference}, & \text{if } difference_t > 60 \\ \textit{medium\_difference}, & \text{if } 30 \geq difference_t > 60 \\ \textit{short\_difference}, & \text{if } difference_t < 30 \end{cases} \tag{5.2}$$

The opportune-moment-identification model is a stochastic environment considering the way of state transitions occur. As presented earlier, S^{OMI} is mainly composed of parameters pertaining to people's varying momentary contexts. Therefore, concerning transitions from state omi_s_t to omi_s_{t+1}, the values of omi_s_{t+1} are observed from people's daily activities. This means that this model does not have a well-defined transition probability matrix.

5.2.1.1 Selective Eligibility Traces

Fig. 5.3 depicts how the environment and agent interact with each other in the opportune-moment-identification-model. a_i denotes the action taken when the environment is in state s_i and r_i is the reward received for a_i. The boxes labeled

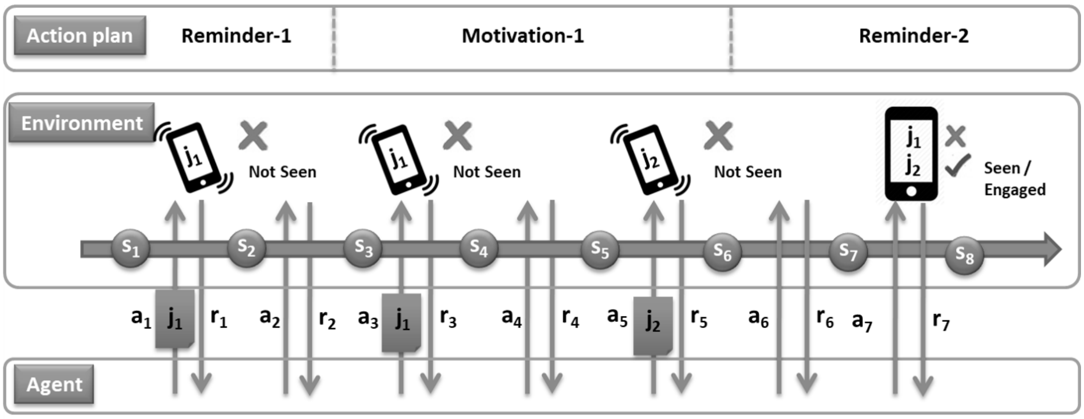

Figure 5.3: Agent-environment interactions during the intervention delivery process

with *j* indicate that an intervention is delivered to the person as an action i.e. *Deliver_Intervention* has been taken; otherwise means that the *Deliver_Nothing* action has been taken. $\forall dp, dp \in DP_{RW}$, for each decision point of the opportune-moment-identification model, the algorithm makes a decision on delivering the intervention (i.e. the one selected by the intervention-selection model) or not at each step based on its internal policy until the behavior is performed or the person engages with the intervention. As seen in the figure, throughout a learning episode, the algorithm might deliver an intervention several times. However, the person might not see or engage with interventions immediately. One or more interventions might be processed or discarded after a certain amount of time by the person when s/he interacts with the delivered interventions. In the example, there would be two delivered interventions, i.e. j_1 and j_2, at state s_7.

Delayed response to the delivered intervention necessitates rewarding past actions. The RL literature has a solution for such cases: eligibility traces. However, the standard eligibility trace approach is not suitable for the proposed approach. Because, for example, assume that the person would only engage with j_2 and discards j_1. r_7 would be a positive reward as the person has seen/engaged with j_2. In this case, the standard eligibility tracing mechanism gives credit to all previous actions, including a_1, a_3 and a_5. However, rewarding a_1 and a_3 is an inconsistency. Because, those actions are suboptimal as interventions were sent in inappropriate moments without any engagement.

This inconsistency is addressed by customizing the eligibility criteria of retrospective rewarding. The proposed approach is basically to assess whether an intervention is engaged or discarded once seen by the person and reward only the engaged ones. Traditionally, all the traces would have been rewarded in the same way in proportion to their weights. For this domain, however, it is obvious that the discarded interventions would not contribute performing the targeted behavior. So, previous actions resulting in discarded interventions are not rewarded. Complementary to this modification, it is ensured that the action just before the engagement with the intervention would be *Deliver_Action* even if it were originally *Deliver_Nothing* action. For example, in Fig. 5.3, although a_7 is a *Deliver_Nothing* action originally, it was indeed an opportune moment for engaging with the intervention. So, the eligibility trace is modified as if a *Deliver_JITAI* action were taken in s_7. This modification lets the algorithm to favor taking *Delivery_JITAI* action in future visits to s_7. The algorithm described in Fig. 5.4 shows the two modifications performed on the standard eligibility trace mechanism.

5.2.1.2 Transfer Learning Across Opportune-Moment-Identification Environments

To be able to apply transfer learning in an RL setup, first, the similarities / differences of the environments among which the knowledge to be shared should be analyzed. Our intention is to apply the transfer learning method in opportune-moment-identification model. In this respect; while the state space, action space and reward functions have the same configuration, only the transition functions differ, which is obvious as they reflect the distinctive characteristics of people.

Having these similarities and differences, it is decided to have complete policies as the knowledge to be transferred to achieve a jump start in terms of effectiveness of the algorithm at the beginning phase of the learning process. Having the policy of others, a learning agent could perform better in unknown states. It may consider choosing the actions taken by the other agents in the same or similar states, instead of choosing a random action. However, this requires identification of other agents, whose policies to be transferred to the current agent. To deal with this problem, a

ALGORITHM 1: Selectively rewarding of eligibility traces

Input:

traces_in_episode (*traces*)	: The list of eligibility traces for the current episode at step t
current_state (s_t)	: The current state
current_action (a_t)	: The last selected action
received_reward (r)	: The last received reward
next_state (s_{t+1})	: The next state upon taking action a_t
next_action (a_{t+1})	: The action to be selected in the next step
RL_policy (Q)	: The internal policy of the RL algorithm
reacted_to_intervention ($reacted_t$)	: Whether the person reacted to the intervention at step t

Constants:

Lambda-factor (λ)	: The lambda factor used for adjusting the retrospective rewarding of eligibility traces
RL_discount_factor (γ)	: The discount factor rewarding the previous actions

Output:

updated_trace_list (*traces*)	: Updated eligibility trace list

▷ update existing traces
for each tr in *traces*
    if tr.opportune_moment = ***true***
        ▷ generate reward through Algorithm 7 by specifying that $reacted_t$parameter is true.
        ▷Assume that the other parameters are provided properly
        r ← generate_reward ($reacted_t$: ***true***)
    $\delta = r + \gamma * Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)$ ▷calculate temporal difference
    tr.q ← tr.q + δ * tr.eligibility   ▷ update q-value of (s,a) referred by the trace
    tr.eligibility = tr.eligibility * λ * γ   ▷update eligibility factor of the trace

▷create new trace
tr ← new Trace(state: s_t, action: a_t, eligibility: λ * γ, opportune_moment: $reacted_t$)
if $reacted_t$= ***true*** and tr.action = *Deliver_Nothing*
    tr.action = *Deliver_Intervention*
traces.add(tr)

Figure 5.4: Modifications on the standard eligibility traces

common policy (CP) is maintained that aggregates all the actions taken by all the agents. The common policy acts as a case repository accumulating the transitions happened in all the environments [92].

Definition 1: (Common Policy - CP): *CP:* $S^{OMI} \rightarrow \{(a, n) \mid a \in A^{OMI}, n \in \mathbb{N}+\}$ where *CP* is a function that maps states to action-number tuples. Each tuple includes the action along with the number of agents taking that action in the given state. CP is not transfered among environments as it is but it is used as a raw data source to train a supervised-learning based *State Classifier (SC).*

Definition 2 (State Classifier - SC): *SC:* $S^{OMI} \rightarrow A^{OMI}$, where *SC* is a function from the state space to the action space, providing classification capability for states that have not been discovered by the current agent. It might even be the case that a state might have not been discovered by any of the agents. Even so, the classifier is able to produce predictions considering the similarity of the states that were previously discovered.

SC is trained during the execution of the overall learning process, which is performed via a set of trials, each of which includes a set of learning episodes. *CP* is updated after each episode, based on the activities of the agent throughout the episode. The update algorithm for CP is explained in Fig. 5.5. Once CP is updated, data items are generated out of it to train SC. Each training data item is a tuple with a state and label as follows: *data_item = (s, l), where* $s \in S^{OMI}$ *and* $l \in A^{OMI}$. The label of each data item is determined by considering the selection counts of actions by the agents. Simply, the action with the highest number of selection for the given state is set as the label.

As presented before, the state parameters of the opportune-moment-identification model are enumerated values, which makes decision trees an appropriate method to perform classification on the data items. Specifically, it has been decided to use the Random Forests [93] as they are more suitable to realize customized transfer learning approaches, such as transferring partial policies, by modifying the tree generation accordingly. In this study, though, the default tree generation and classification mech-

```
ALGORITHM 2: Updating Common Policy
Input:
    list_of_state_action_counts (CP)      : List of state-action counts aggregated so far, i.e. the
                                            current Common Policy (CP)
    episode_analysis (ea)                 : Episode analysis keeping the history of visited states, taken
                                            actions and received rewards throughout the episode
    Output
▷update the state-action counts
for each ea_item ea.state_list            ▷ea.state_list contains elements that include the states
                                            visited, action taken, and reward received (st , at , rt)
                                            throughout the episode
    count ← CP [ea_item.s][ea_item.a]     ▷current count of selection of action a in state s
    count ← count + 1
    CP [ea_item.s][ea_item.a]← count
▷update the labels of states
for each sa_counts in CP                  ▷ sa_countsincludes the counts for all possible action for a
                                          specific state
    max_count ← 0
    for each sa_count in sa_counts
        if sa_count > max_count then
            max_count ← sa_count
            sa_counts.label ← sa_count.a    ▷ assign the action with the highest count as the label of the
                                              state represented in the current sa_counts
```

Figure 5.5: The algorithm for updating CP after each episode

anism provided by the Apache Spark library[94] is used. By default, a random forest grows many classification trees such that each tree is trained with a subset of the initial training data. To classify a new data item, the item is put down each of the trees in the forest. Each tree gives a classification output and the forest chooses the classification having the higher output over all the trees in the forest.

5.2.2 The Intervention-Selection Model

The MDP for the intervention-selection model is represented with: $M^{IS} = <S^{IS}, A^{IS}, P^{IS}, R^{IS}>$.

S^{IS} is the state set, where each state s^{IS} such that $s^{IS} \in S^{IS}$ is represented with a tuple composed of four parameters as follows:

s^{IS} = <*habit_strength, behavior_frequency, day_type, remember_behavior*>

Having the same reasoning with the opportune-moment-identification model concerning not having large state and action sets, all the state parameters are modeled as discrete parameters as follows:

- ***habit_strength*** = *{x: x >= 0 and x <10, x* $\in \mathbb{N}$*}.* Habit strength represents the strength of automaticity of the behavior without occupying the mind to remember performing the behavior. In other words, the higher the habit strength is, the less need for providing cues in the form of interventions for reminding the behavior.

- ***behavior_frequency*** = *{x: x >= 0 and x <10, x* $\in \mathbb{N}$*}.* As described previously, behavior frequency is the ratio of number of times executing the behavior to the total number of opportunities to perform a behavior, within a certain time frame.

- ***day_type:*** Day type is an enumeration that can take either *weekend* or *weekday* values. Its value is extracted from the date of the active learning episode i.e. the simulated day.

- ***remember_behavior:*** Remembering behavior is also an enumeration that can take either *true* or *false* values. The specific value of this parameter indicates whether the person would remember performing the behavior as planned in the action plan.

The action set of the intervention selection model is defined as follows:
$A^{IS} = \{J_1, J_2, \ldots, J_n, No\_Intervention\}$

The set includes intervention types, i.e. $J_1, \ldots, J_n$ that are supposed to be identified by domain experts addressing a particular disease/health problem. In addition to the specific intervention types, not delivering any intervention, represented with *No_Intervention* label, is also a possible action. However, not all actions are available for each decision point but only the interventions included in the EA_{dp} are eligible at each decision point.

R^{IS} function is conditioned on remembering the behavior at state s_t, delivering an intervention and reacting to the delivered intervention. At each time step *t* during the learning process, the reward is calculated based on these variables. Varying combinations of conditions on these variables form the reward function as presented in Eq.

5.3. The values for the six respective cases in the equations are instantiated as *-3, -10, 10, -5, -1* and *-50*. The reasoning behind such an instantiation is to generate relatively higher rewards in cases when the person does not remember performing the behavior and an intervention type is selected as a counter action. Engaging with the intervention is rewarded positively so that the interventions that are not engaged by the person would be selected less. The worst decision of the algorithm would be not to select any intervention even if the person would not remember performing the behavior. So, this is the case where the largest reward, a negative one, is generated.

$$
is\_r_t = \begin{cases}
\textit{sent\_remembered\_reacted}, \text{ if} & \begin{array}{l} is\_a_t \neq No\_Intervention, \\ remember\_behavior(s_t) = true, \\ reacted_t = true \end{array} \\[2em]
\textit{sent\_remembered\_not\_reacted}, \text{ if} & \begin{array}{l} is\_a_t \neq No\_Intervention, \\ remember\_behavior(is\_s_t) = true, \\ reacted_t = false \end{array} \\[2em]
\textit{sent\_not\_remembered\_reacted}, \text{ if} & \begin{array}{l} is\_a_t \neq No\_Intervention, \\ remember\_behavior(is\_s_t) = false, \\ reacted_t = true \end{array} \\[2em]
\textit{sent\_not\_remembered\_not\_reacted}, \text{ if} & \begin{array}{l} is\_a_t \neq No\_Intervention, \\ remember\_behavior(is\_s_t) = false, \\ reacted_t = false \end{array} \\[2em]
\textit{not\_sent\_remembered}, \text{ if} & \begin{array}{l} is\_a_t = No\_Intervention, \\ remember\_behavior(is\_s_t) = true \end{array} \\[2em]
\textit{not\_sent\_not\_remembered}, \text{ if} & \begin{array}{l} is\_a_t = No\_Intervention, \\ remember\_behavior(is\_s_t) = true \end{array}
\end{cases}
\tag{5.3}
$$

P^IS is a matrix of probabilities, $P(is\_s_{t+1}|is\_s, is\_a_t)$, indicating the probability of transition from state is_s to state is_s_{t+1} by taking the action is_a_t. The intervention-selection model follows an episodic learning approach such that each day corresponds to an episode. State transitions throughout an episode, i.e. a day, is laid out by the action plan associated to the person. As described below, the transition dynamics from state is_s_t to is_s_{t+1} are deterministic. In the rest of this section, transition dynamics about the individual state parameters are presented.

The subsequent value of the *day_type* parameter, i.e. the value for the state *s'*, is trivial. It is simply extracted from the date associated with the simulated day. The dynamics of the *habit_strength* and *remember_behavior* parameters are determined by the habit formation model that are adopted from [84]. Similar to the RL models, the habit formation model is also an iterative model and has its own internal dynamics. It runs in parallel with the intervention-selection model. At each state transition of the RL model, the habit formation model is also iterated by updating its internal parameters including habit strength and behavior frequency. The time-series data generated by the habit formation model is defined with the set $HFTS$ as introduced in Sec. 3.4.

Four parameters are updated at each step of the habit formation model, namely *habit strength, behavior frequency, accessibility* and *salience of reminders* (interventions in our case). Although accessibility and salience of reminder concepts are not directly included in the laid-out RL model, they are essential parameters for predicting the remembering of the behavior. In the rest of this section, the updating procedures of these four parameters are presented via a set of equations that are adopted from the habit-formation model. Following the update procedures, Eq. 5.13 and Eq. 5.14 show how the prediction on remembering the behavior is made. All these equations are solved at each step of the intervention-selection model in order to calculate the parameters of the next state.

The following three equations show the calculation of the next *habit strength* value (hs_{t+1}). In Eq. 5.4, the decay amount on the habit strength arising from not performing the behavior ($bp_t = false$) is obtained. If the person performs the behavior ($bp_t = true$) the habit strength does not decrease. Then, in Eq. 5.5, the amount of increase on the habit strength is calculated in case the person has performed the behavior. Finally, the final value of the habit strength for the next step *(hs_{t+1})* is obtained in Eq. 5.6 by subtracting the decay amount from and adding the increase amount to the value at the previous step *(hs_t)*.

$$habit\_strength\_decay_t = \begin{cases} hs_t * HDC, \text{if} & bp_t = false \\ 0 & \text{otherwise} \end{cases} \tag{5.4}$$

$$habit\_gain_t = \begin{cases} (hs_t * (1 - bf_t) * HDC, \text{if} & bp_t = false \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

$$hs_{t+1} = hs_t - habit\_strength\_decay_t + habit\_gain_t \tag{5.6}$$

In these equations, hs_t, bf_t and bp_t are the habit strength, behavior frequency and behavior performance at step *t*. bf_t indicates whether the behavior is performed or not after the action a_t. ***HDC***, i.e. the habit decay constant, is the constant introduced by the habit formation model, adjusting the decay amount at each step. (Note that all the habit-related constants in the equations in this section are obtained from the habit formation model.)

Eq. 5.7 shows the calculation of the next value of *behavior_frequency* (bf_{t+1}). A sliding window is maintained to keep a record for each opportunity to perform the behavior within a particular period in the past. As a result of collaborative discussion with cognitive behavior psychologists, it is decided to have the history size as *14 days*, as records of past 2 weeks might contain certain patterns about the behaviors of people about their self-management activities. The window is implemented as a queue such that each of its items is a binary value indicating whether the behavior has been performed or not. At each step, the oldest item (bp_{oldest}) of the window is removed from the queue, and the latest behavior performance indicator (bp_{newest}) is pushed into the queue. The behavior frequency is the ratio of the number of the behavior performances to the total opportunity count, as calculated in Eq. 5.7.

$$bf_{t+1} = (bf_t * opportunity\_count - bp_{oldest} + bp_{newest})/opporunity\_count \tag{5.7}$$

People get bored seeing the same type of notifications on their mobile phones. They become unresponsive to such notifications or discard them [95]. In the scope of this study, intervention types are grouped according to their categories (e.g. reminders or motivations as mentioned earlier). Several intervention types with different configurations e.g. underlying psychological methods might be related with the same intervention category. For example, two reminder intervention types might be using different behavior change techniques to remind a behavior. Therefore, a distinct salience indicator is maintained for each intervention type. Eq. 5.8 shows the calculation of the next values of intervention saliences ($sal_{i,t+1}$). Salience of interventions, with the same category with the selected intervention ($category_i = category_{sa_t}$) are updated at each step.

The salience of the selected intervention type (is_a_t) decreases in proportion to the current salience amount. The amount is calculated by using the *SC* constant of the habit formation model. The salience values of other interventions having the same category with the current intervention ($i \neq is\_a_t$) are increased as the person is not exposed to them. The same constant is used to calculate the decrease amount by applying the opposite calculation on the current salience amount. Eq. 5.8 implements the narrated logic. In the equation, the *category* concept is a function returning the category of a given action.

$$
\begin{aligned}
&\forall_i (i \in A^{IS} \ and \ i \neq \textit{No\_Intervention} \ and \ category(i) = category(is\_a_t)),\\
&sal_{i,t+1} = \begin{cases} sal_{i,t} - sal_{i,t} * SC, \text{if} & i = is\_a_t \\ sal_{i,t} + sal_{i,t} * SC, \text{if} & i \neq is\_a_t \end{cases}
\end{aligned}
\tag{5.8}
$$

According to the habit-formation model, the accessibility decreases at each step, in proportion to its current amount (Eq. 5.9). On the other hand, it increases in presence of a reminder (Eq. 5.10) and when the targeted behavior is performed (Eq. 5.11). In our case, interventions are the reminders increasing the accessibility value. So, if an intervention is selected as the action at step *t*, the accessibility is increased accordingly. The updated value of the accessibility is obtained by subtracting the decay amount from and adding the increase amounts to the previous value (Eq. 5.12).

$$acc\_decay_t = acc_t * ADC \tag{5.9}$$

$$acc\_gain\_rem_t = \begin{cases} \begin{array}{l}(AGC\_R + (1 - AGC\_R) \\ \quad *WCI\_REM * ci) * sal_t\end{array} \text{if} & is\_a_t \neq No\_Intervention \\ 0 & \text{otherwise} \end{cases} \tag{5.10}$$

$$acc\_gain\_pb_t = \begin{cases} bf_t * AGC\_PB, \text{if} & bp_t = true \\ 0 & \text{otherwise} \end{cases} \tag{5.11}$$

$$acc_{t+1} = acc_t - acc\_decay_t + acc\_gain\_pb_t + acc\_gain\_rem_t \tag{5.12}$$

In these equations, $acc_t, bp_t, bf_t, is\_a_t$ and sal_t represent accessibility value, whether the behavior performed, behavior frequency, selected action in the intervention-selection model and salience of the selected intervention at step t. ci represent the commitment intensity of the person indicating the strength of any form of internal pressure by the person. Furthermore, ADC, AGC_PB, AGC_R and WCI_REM are the constants of obtained from the habit formation model. They respectively represent the constants for accessibility decay, accessibility gain constant by performing the behavior, accessibility gain by reminders and weight of the commitment intensity in accessibility gain by reminders. For more details on the model, please refer to the original study [84].

The last step regarding the iteration of the habit formation model is to predict whether the person would remember performing the behavior or not. The prediction is simply done by comparing the current accessibility value and the accessibility threshold value. Accessibility threshold depends on the habit strength and behavior frequency parameters. Based on these values, the threshold value is obtained using Eq. 5.13. Getting the accessibility greater than the threshold indicates that the person would remember performing the behavior (Eq. 5.14).

$$acc\_threshold_t = \begin{array}{l} ATC - (ATC * ATH * hs_t) \\ +(1 - ATC) * ATBF * bf_t * (1 - DTH * hs_t) \end{array} \tag{5.13}$$

$$behavior\_prediction_{t+1} = \begin{cases} true, \text{if} & acc_t \geq acc\_threshold_t \\ false & \text{otherwise} \end{cases} \tag{5.14}$$

In this equation, acc_t, hs_t and bf_t represent accessibility, habit strenght and behavior frequency values at step t. $ATC, ATH, ATBF$ and DTH are the constants representing accessibility threshold, weight of habit strength in accessibility threshold, weight of the behavior frequency in accessability threshold and distraction weight due to habits. Similarly, to have better insight about the habit formation related concepts, please refer to the original model. Presenting how the remembering is predicted concludes the transition dynamics of intervention-selection model.

5.3 Overall Algorithm

In this section, the overall learning algorithm, combining all the pieces presented so far, is described. The algorithm is partitioned into a set of sections for the sake of clarity and easy understanding. Each learning episode has a daily time frame. In other words, each day of people's lives is processed throughout an episode. The overall algorithm, in this respect, describes the procedure followed throughout a single learning episode.

The first section shows the inputs, all of which are complex data structures including specific data elements used in the algorithm. The first four inputs are main components of the RL environments, namely *environment* and *agent* objects of intervention-selection and opportune-moment-identification models. The environment objects keep track of the current state of the environment and transition history. The agent objects are a proxy for the learning policy, which is used to select actions. The next input is the *habit_formation_model*, which keeps the variables like habit strength or behavior frequency required for simulating the habit formation model. As described earlier, an *action plan* includes daily planned activities for the person. Lastly, the *State Classifier* is the trained model that is used to predict actions to be taken in unknown states.

The actual procedure starts with the second section, which in turn starts with the outermost loop of the algorithm. This loop is executed such that $\forall dp, dp \in DP_{AP}$, for each decision point of the intervention-selection model throughout a learning episode. The first step is to determine the action *(is_a_t)* to be taken in the current state *(is_s_t)* of the intervention-selection model, by the learning agent of the intervention-selection model. Internally, a greedy policy is used such that the action leading to the highest long-term reward is selected. However, the policy does not select the action among the complete set of intervention initially configured but from the set of eligible actions for the decision points, i.e. EA_{dp}.

In the third section, the opportune-moment-identification model is run only if a certain intervention type is selected ($is\_a_t \neq No\_Intervention$). The first operation performed by the opportune-moment-identification is to determine the action *(omi_a_t)* based on the current state *(omi_s_t)* of the opportune-moment-identification model. The same greedy logic is used by this model also. However, the learning algorithm utilizes *SC* in case it has to select a random action ($omi\_a_s.selectMode = RANDOM$). Such cases include encountering with an unknown state or having same q-values for all actions in a state. In other cases, it utilizes the greedy policy and selects the action with the highest q-value. Once the action selection is finalized, the environment makes transition from the current state to the next state by taking the selected action. For the opportune moment identification model, this basically means an update in the person's momentary context.

In the fourth section, two simulations take place. If the intervention is delivered ($omi\_a_t = Deliver\_Intervention$), the person's reaction to the delivered intervention is simulated. The result of the simulation is either to discard the intervention or engage with it. The second simulation is for performing the targeted behavior. Whether the performance will be performed or not is determined by the habit formation model. Details about the two simulation activities will be given in the next section.

The fifth section starts with obtaining the reward *(omi_r_t)* for the action taken by the opportune-moment-identification model via Eq. 5.1 and Eq. 5.2. Then, eligibility traces are updated via $Algorithm\ 1$, and the transition is recorded into the episode analysis object. At each step of the opportune-moment-identification model, the current time is advanced according to the simulated daily activities. The decision making on delivering the selected intervention lasts, in other words an episode of opportune-moment-identification model runs, as long as the current time is within the time frame associated with the current planned activity of the action plan ($context\_change$ occurs within $action\_plan\_activity.time\_frame$). Once the opportune-moment-identification episode is over, CP is updated with the collected data during the episode via $Algorithm\ 2$.

Following the fifth section, execution context switches back to the intervention-selection model by advancing the habit formation model one step. During this phase, remembering the behavior parameter, habit strength, behavior frequency, salience of interventions and accessibility of the behavior are updated as elaborated earlier by evaluation of habit formation related equations from Eq. 5.4 to Eq. 5.14.

Finally, the next step for the intervention-selection model (is_s_{t+1}) is obtained using the updated parameters of the habit formation model and the reward (is_r_t) is generated via Eq. 5.3.

```
ALGORITHM 3: Learning with intervention-selection and opportune-moment-identification
Input:
    intervention_selection_env (is_env)
    intervention_selection_learning_agent (is_ag)
    opportune_moment_identification_env (omi_env)                                       Section 1
    opportune_moment_identification_learning_agent (omi_ag)
    habit_formation_model (hfm)
    action_plan (AP)
    common_policy (CP)
    state_classifier (SC)
▷ JITAI-selection
for each decision point dp (at step t) in DP_AP                                         Section 2
    is_s_t ← js_env.current_state
    is_a_t ← js_ag.learning_policy.getAction(is_s_t)  ▷ action is selected from EA_dp

    ▷ perform the activities within the timeframe of the current activityin the opportune moment
    ▷ identification environment
    if is_a_t ≠ No_Intervention then
        ▷ find the action plan activity from AP from which the decision point is generated
        action_plan_activity ← getActivityForDecisionPoint(dp)
        for each context_change (at step s) in action_plan_activity.time_frame
            omi_s_s ← omi_env.current_state                                             Section 3
            omi_a_s ← omi_ag.learning_policy.getAction(omi_s_s)

            ▷ if the action was selected randomly then consult to state classifier for a better decision
            if omi_a_s.select_mode = RANDOM then
                omi_a_s ← SC.guessAction(omi_s_s)
            ▷get next state and generate reward for the opportune moment identification model
                omi_s_s+1 ← omi_env.getNextState(omi_s_s, omi_a_s)

            ▷ simulate reaction to intervention and behavior performance
            if omi_a_s = Deliver_Intervention                                           Section 4
                reacted_to_intervention ← simulate_reaction_to_intervention()
            performed_behavior_t ← simulate_behavior_performance(hfm.remember_behavior)
            ▷ generate reward for the opportune moment identification environment
            omi_r_s ← calculate_omi_reward (omi_a_s, differences_s,
                                              reacted_to_intervention) ▷ Eq. 5.1 & Eq. 5.2
            update_eligibility_traces (omi_env. traces, omi_s_s,omi_a_s, omi_r_s, omi_s_s+1,
                                 omi_ag.learning_policy.getAction (omi_s_s+1),
                                 omi_ag.learning_policy, reacted_to_intervention)    ▷ Algorithm 1   Section 5
            omi_env.episode_analysis.add_record(omi_s_s, omi_a_s,omi_r_s)

        ▷ update the common policy when the episode ends
        update_state_classifier (omi_env.recorded_episodes, CP)                      ▷ Algorithm 2
    ▷ execute a step in the habit formation model
    update_habit_strength (hfm.habit_strenght_t, hfm.behavior_frequency_t,
                                  performed_behavior_t) ▷ Eq. 5.4 & Eq. 5.5 & Eq. 5.6
    update_behavior_frequency (hfm.behavior_frequency_t, is_a_t)                      ▷ Eq. 5.7
    update_saliences (is_a_t, hfm.saliences_t)                                        ▷ Eq. 5.8     Section 6
    update_accessability (hfm.accessability_t, hfm.remember_behavior_t, hfm.behavior_frequency_t,
                   is_a_t, hfm.saliences_t)          ▷ Eq. 5.9 & Eq. 5.10 & Eq. 5.11 & Eq. 5.12
    hfm.remember_behavior_t+1 ← predict_behavior (hfm.accessability_t, hfm.habit_strength_t,
                          hfm.behavior_frequency_t)                  ▷ Eq. 5.13 & Eq. 5.14
    ▷get next state and generate reward for the intervention selection mode
    is_s_t+1 ← js_env.getNextState(js_s_t, js_a_t)
    is_r_t ← generate_reward_for_intervention_selection (is_s_t, is_a_t, is_s_t+1,       Section 7
                                               reacted_to_intervention) ▷ Eq. 5.3
```

Figure 5.6: Overall learning algorithm

CHAPTER 6

COMMUNICATION ENGINE: THE SOFTWARE REALIZING THE JITAI DESIGN AND DELIVERY PLANNING

Communication Engine is the implementation of the conceptual approach described so far. Below, technical details about it are presented by elaborating the architectural design, individual components and interactions between the components.

6.1 Reactive Programming within Lambda Architecture

A *person context* contains dynamic parameters e.g. physical activity status, location, blood glucose measurements of which values are continuously updated according to streaming data in real-time. Communication Engine ingests such streams of data, processes and uses them to run the intervention delivery mechanism. In this respect, Communication Engine is an event-driven application reacting to external events. So, the *reactive programming*, which is a well-suited paradigm for working on asynchronous data streams[96], is adopted to implement such a system. The following technologies has been used to realize various reactive programming tasks:

- *Apache Kafka*[1] for establishing a publish/subscribe mechanism between the data sources and Communication Engine to monitor the new observations related to persons and actions performed by the persons.
- *Apache Spark*[2] and *Spark Streaming*[3] for distributed processing of incoming data streams as well as intervention delivery planning.

[1] https://kafka.apache.org/
[2] https://spark.apache.org/
[3] https://spark.apache.org/streaming/

• *Apache Cassandra*[4] for persisting the person state, which contains all the information required to run the learning algorithm.

Considering the need for continuous monitoring of people's changing context, performing data management tasks in frequent intervals, running learning algorithms on the recent context data and performing all these operations for all persons a scalable architecture is proposed. The architecture can scale with increased number of participants and data sources. In general, the implementation approach follows the lambda architecture which is a scalable data-processing architecture designed to handle massive quantities of data by taking advantage of both batch and stream-processing methods [97] as illustrated in Fig. 6.1.

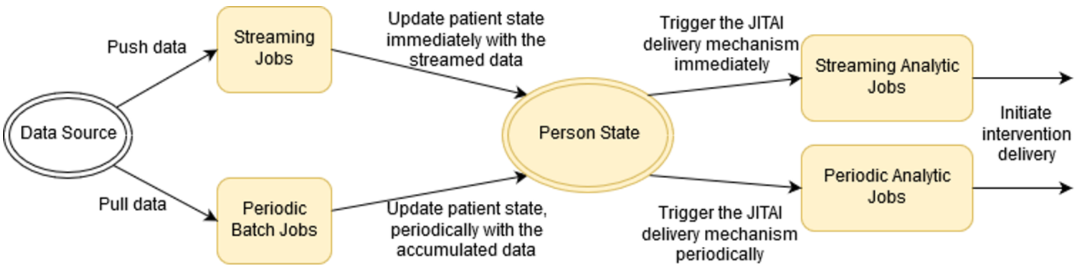

Figure 6.1: Data processing architecture of Communication Engine

Data is processed in two layers namely, batch layer and speed (streaming) layer, which is the main characteristics of the lambda architecture. In the batch layer, daily jobs run processing the data of all persons and update their person state. Batch jobs are executed to retrieve the data that is not changed frequently like goals and action plans. Streaming jobs handle the data that need continuous monitoring to update patient state and react immediately on changes.

The two data processing layers are also applicable to trigger the intervention delivery mechanism. In the batch layer, the mechanism is triggered periodically for a specific period e.g. every 10-minutes. The streaming layer reacts to significant changes on the data streams and trigger the mechanism with the updated person state.

[4] http://cassandra.apache.org/

6.2 Component Architecture

The reactive programming and lambda architecture paradigms are realized by several components interacting with each other as depicted in Fig. 6.2. The entry component in the architecture is the *FHIR repository*, which persists the data generated automatically by devices / sensors or manually by persons themselves in *HL7 FHIR* format[5]. The repository exposes the managed data via a REST API for on demand requests or pushes the continuous stream data to Kafka. *Stream Manager* keeps track of the registered data channels and acts as an abstraction layer on the channels by transforming the streamed data into proper format so that the rest of the components process it in a distributed manner.

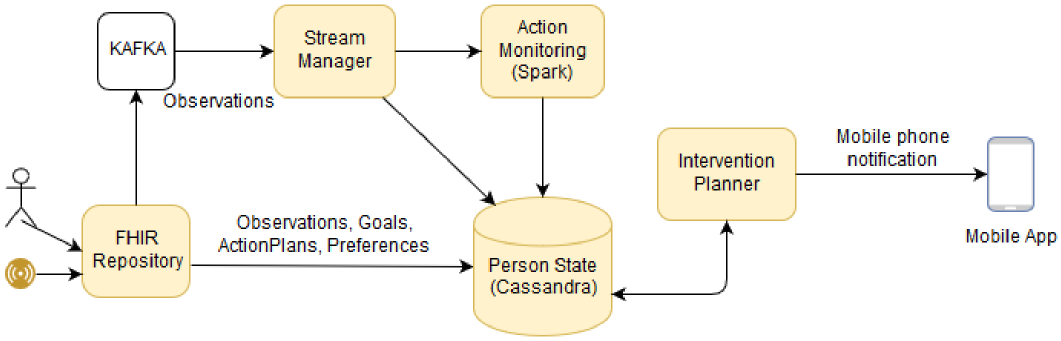

Figure 6.2: Internal architecture of Communication Engine

Person State contains any information required during the intervention planning and delivery process. This component provides the *Action Monitoring* and *Intervention Handler* components with up-to-date person state when requested. Vice versa, this component keeps the person state always up-to-date both with the raw data retrieved from the FHIR repository and with the deduced data generated by Action Monitoring and Intervention Handler components.

A person state is composed of quite diverse information including scheduled activities driven from the action plans, active goal, activities performed, delivered intervention related to the scheduled activity and a set of behavioral and health metrics that might represent a single measurement like blood glucose value at a specific time or a calculated value as a result of an analytic procedure e.g. average blood glucose values

[5] https://www.hl7.org/fhir/

during the last week, percentage of abnormal blood glucose measurements or 3 highest readings in last 7 days. These variables contain the tailoring variables composing the RL states. It is possible to expand the patient state with additional metrics without disrupting the rest of the system. Employing Cassandra enables distributed storage of the aggregated data within person states.

Action Monitoring runs as a Spark Streaming Job by processing the streams of person metrics including the latest observations coming from the FHIR repository such as blood glucose measurements, physical activity logs, dietary intake logs, medication intake logs, etc. It then tries to match the observations related to a person with the scheduled activities for him/her. Once such a match is detected, the person state is updated with the information indicating how well the person adhered to the scheduled activity in terms of timing, intensity and overall performance. For example, a new blood glucose measurement will mark a scheduled blood glucose monitoring action as performed along with the adherence performance.

Intervention Planner performs a set of tasks for evaluation of the conditions to initiate an intervention, preparation of its content and its delivery to the person. Intervention Planner itself is a complex component with sub-components specialized for distinct tasks as depicted in Fig. 6.3. It acquires the activities specified in the action plan of the person at a daily basis. Within the time frame associated with the action plan activity, *Action Plan Monitor* periodically monitors the person state to check whether the person has performed the activity or not. In either cases, Action Plan Monitor triggers the *Intervention Decision Handler* in the valid time frame. Intervention Decision Handler, in turn executes the RL algorithm to decide whether an intervention will be delivered to the person or not. Before delivering the intervention, *Performance Analyzer* finalizes the intervention by calculating the placeholders included in the content.

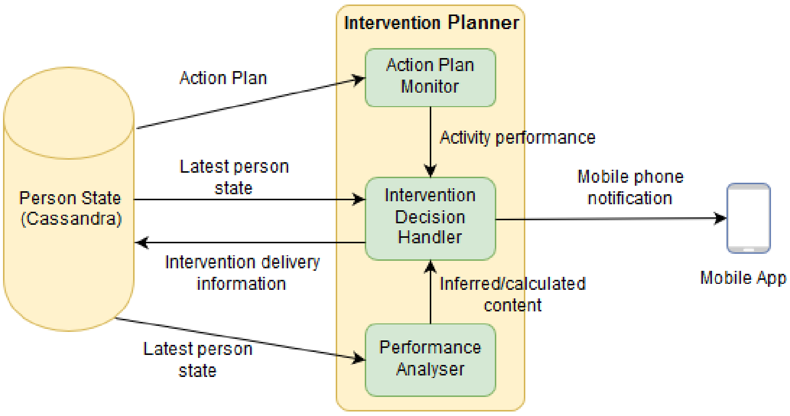

Figure 6.3: Internal architecture of Communication Engine

CHAPTER 7

VALIDATION

The validation activities are two-folded. Firstly, how the expandable JITAI design mechanism meets the JITAI design-related specifications derived from various resources like BCT taxonomies, clinical guidelines or algorithms for automated self-management support is shown. Then, a simulation-based case study is laid out to validate the proposed algorithm with respect to the adaptation of the intervention delivery strategy according to a set of simulation parameters. The scope of this study is limited in terms of application of the learning algorithm in real-life case studies. The main aim is to validate the algorithm in simulated settings before deploying it in a real-life case study involving human participants. Nevertheless, a small-scale real-life case study, utilizing a preliminary version of the algorithm, is also presented.

7.1 Validating the JITAI Design Mechanism

Validation of the JITAI-design mechanism is mainly done by relating the relevant parts of BCTs, clinical guidelines and self-management support algorithms with the proposed JITAI design mechanism. The aim is to identify the requirements to devise an automated, theory-driven self-management support enabled by the integration of the JITAI design mechanism and the learning algorithm. For example, a BCT taxonomy provides alternatives ways of motivating a person, which indicates a relation to the intervention option element of JITAIs. Clinical guidelines recommend certain amounts of a certain activity to performed within a certain amount of time. Considering the duration and amount criteria, an intervention should be associated with proper decision rules executed at appropriate decision points, which implicitly requires integration of relevant data sources as tailoring variables. This pattern applies also to any effort aiming to automate self-management support. Below, after providing examples about these concepts, it is argued that the proposed approach is capable of meeting the JITAI-related specifications included in the existing resources.

Relating BCTs with the JITAI design mechanism: CALO-RE is a taxonomy introducing 40 types of BCTs aiming to change people's behaviors related to physical activity and eating in a healthier manner[88]. The taxonomy guides prospective implementers about the content of BCTs without any other specific details. As the motivation is to provide automated self-management support, BCTs are evaluated with respect to the possibility to automate them by conditioning on a set of input parameters. This is a qualitative approach through which the suitability of each BCT is evaluated for automation based on its description in the taxonomy. 31 out of 40 BCTs introduced by CALO-RE are classified as suitable for automation. *Goal setting, action planning, fear arousal* and *prompting rewards contingent on effort or progress towards behavior* are some of the BCTs that are classified as suitable for automation.

BCTs that are not evaluated as suitable for automation are mostly the techniques that require physical meeting of the care receivers and care givers, e.g. *motivation interviewing* or *agreeing on behavioral contracts*. Some other BCTs are hard to be automated as management of the relevant data is not feasible. For example, *prompting generalization of a target behavior* and *environmental restructuring* require representing, retrieval and processing the information about the physical environment in which the behavior is performed. Appendix B includes the complete list of BCTs that are classified as suitable or not suitable for automation. Furthermore, Appendix C shows two example JITAI definitions related to two BCTs availabile in the CALO-RE taxonomy. Overall, it is argued that the techniques classified as suitable can be realized with the proposed design mechanism.

Relating clinical guidelines, algorithms and available studies with JITAI design: Clinical guidelines provide clear starting points for JITAI components, even though the relation between the JITAI framework and those resources is not stated explicitly. American Diabetes Association (ADA)[98], the Joslin Clinical Guideline for Adults with Diabetes[99] provide recommendations specific to diabetes disease. Some of the recommendations given by these guidelines are about people's life-style related to physical exercise, nutritional intake or self-monitoring of blood glucose. For example, Joslin guideline recommends blood glucose monitoring 4-6 times per day, 2-4 consecutive days of postprandial monitoring, 60-90 minutes of physical activity 6-7 days per week. ADA recommends engagement in 60 minutes per day or more physical activity, interrupting prolonged sitting every 30 minutes with short bouts of physical activity. Such life-style recommendations are usually based on the conditions of the person. For example, physical capacity of the person is measured with physical functioning tests[100] before adjusting the intensity of the physical exercise.

Studies aiming to improve self-management of chronic diseases integrate various information manually through quizzes, questionnaires, likert scales about nutritional intake, psychological status, physical functioning capability or automatically through sensors, devices, mobile app usage statistics into their decision-making processes [33, 74]. There are numerous examples on this but as a simple example, the ULTEMAT platform provides an intervention to learn the mood of the person via a likert scale if the person has not specified his/her mood the day before.

In addition to the life-style related interventions, the proposed mechanism can also be used to design interventions aiming to automate clinical decision making to a certain extent. For example, Predictive 303 algorithm introduces a rule set for adjusting insulin detemir every 3 days based on the mean of three adjusted fasting plasma glucose (aFPG) as follows: if mean aFPG < 80 mg/dl, reduce dose by 3 unit; if aFPG is between 80 and 110 mg/dl, no change; and if aFPG > 110 mg/dl, increase dose by 3 unit[101].

Even though the relation between the JITAI framework and the specifications mentioned above is not stated explicitly, the examples exemplify different cases of tailoring variables, decision points as well as decision rules specified in different resources. For example, considering the interruption of inactivity example, the tailoring variable is the physical activity status, the decision point is every minute and the decision rule would be to check whether the person has been inactive for the last 30 minutes. Appendix D includes intervention definition examples based on the JITAI specifications extracted from the aforementioned resources.

Generic requirements to automate self-management support: The specifications presented above lead the following conclusion: Designing a JITAI first requires the integration of the data that would be used in decision-making processes related to intervention delivery. In other words, the required tailoring variables should be in place. Data to be integrated vary in nature (structured / unstructured, static / streaming) and in source (devices, sensors, phone operating system, mobile app, forms or questionnaires). To be able to make reasoning on the integrated data it must be transformed into a format that could be used in a decision rule. As in the interrupting inactivity example, this could be reducing the last 30 minutes of physical activity data to a binary indicator.

As stated in Sec. 4, the JITAI design mechanism is easily expandable in terms of integration of additional data sources, either static or streaming and already provides built-in tailoring variables that process the incoming data. The outcome of a tailoring variable is considered as a metric and the person state could be expanded with as much metrics as desired. Once the person state includes the information in the desired format, decision rules can utilize that information. Decision rules are the elements evaluating the suitability of the conditions to deliver an intervention automatically. Thus, it is argued that the expandable data integration and processing capabilities described above capture the JITAI-design-related requirements driven from existing resources.

Validating the JITAI design capabilities in real-world case study: The JITAI design capabilities have also been validated in the POWER2DM Project. The objective of POWER2DM is to develop and validate a personalized self-management support system for diabetes patients. In the project, computer scientists, behavioral psychologists and internists have collaboratively come up with a set of JITAIs addressing diabetes patients on different behaviors including blood glucose monitoring, exercising, medication adherence and carbohydrate monitoring. Appendix E presents example JITAI definitions in relation to these behaviors.

7.2 Simulated Case Study

A simulation testbed for simulating the treatment and self-management process is laid out to validate the personalization of intervention delivery strategies. Considering the domain of the likely real-life case study, the simulation concepts are related to the diabetes disease, if not generic. The testbed is composed of three main concepts to be simulated namely an *action plan, JITAIs* and *personas*. Care process simulations are performed for four personas with differentiating conditions related to the simulated concepts. The aim is to show that the JITAI personalization algorithm captures the persona-specific conditions and adapts the intervention delivery accordingly. Details about the simulated concepts are in the subsequent sections.

7.2.1 An Imaginary Action Plan

A simple action plan is introduced driven by the targeted real-life case study. In this respect, the aim is to make the diabetes patients form habit for measuring blood glucose levels repetitively. Specifically, the action plan includes only a single activity, which is blood glucose measurement that should be performed before each meal during the day. This means that it is expected that blood glucose measurements will be performed 3 times a day throughout the simulated care process.

A fixed intervention delivery strategy: Driven by this action plan, a fixed intervention delivery strategy is defined as a baseline algorithm against which the results of the proposed approach will be compared. Since the study proposes improvements in the scope of the opportune-moment-identification model only, the fixed intervention delivery strategy has implications only on this model. Specifically, the intervention-selection model selects interventions as introduced before. That is it employs the RL-based learning mechanism. On the other hand, concerning the opportune-moment-identification model, delivery times of the interventions are always fixed in the following way: Interventions are sent at 9:00, 13:00 and 20:00 o'clocks only if a person has not engaged with the intervention and has not performed the behavior yet. Considering the way the opportune-moment-identification works, this means that the *Deliver_Intervention* action is consideredat each decision point until the engagement happens or behavior performed.

7.2.2 An Imaginary Set of Interventions Identified for the Targeted Behavior

It is expected that intervention designers / behavior scientists would configure a set of interventions to achieve the expected outcomes of the care program. The algorithms is supposed to learn personal preferences on these interventions. In this respect, three types of intervention are configured such that two of them are used for reminding the behavior and the third one is used for motivation. Therefore, only the reminder interventions are applicable in a state where reminding the behavior is targeted. On the other hand, only one motivational intervention is applicable where the targeted support is about motivating. Please refer to Fig. 5.1 and Fig. 3.1 that show how interventions with different categories are distinguished. Targeting the blood glucose monitoring behavior, three JITAI instances, each of which implements a specific behavioral change technique (BCT) as described in the CALO-RE taxonomy, are defined. Below, distinguishing characteristics of these three interventions are presented with respecto to their JITAI components. Furthermore, Appendix F shows their instantiation via the proposed design mechanism.

- ***Intervention-1*** Prompting self-monitoring of behavior: The first intervention is a standard reminder. Having the reminder category, the intervention is supposed to be sent within the period during which the activity is supposed to be performed.
- ***Intervention-2*** Reminding with comparing with others: This is also a reminder intervention. While reminding the activity, it also motivates the person by presenting a comparison with others in terms of performance of the targeted behavior. This intervention has the same decision rules with the previous one. So, both will be considered as eligible interventions at the same decision points.
- ***Intervention-3*** Praising the performed behavior: This one is a motivational intervention complimenting the person on successful performance of the planned behavior. It is associated with three decision rules representing the achievement of the monitored goal in daily, weekly and monthly timescales respectively.

Meeting only one of them is sufficient for making this intervention eligible for delivery.

7.2.3 Persona Simulation

Four distinguishing dimensions are considered for the simulated people. The first one is the *habit formation*, which is an indicator of the strength of automaticity of performing the targeted behavior without external intervention. As described in detail in Sec. 5, an evidence-base habit formation model has been integrated to simulate habit-related concepts in a theoretically valid way. According to the model, forming habit takes varying durations for different people as the model is instantiated with person-specific values. The second dimension is *daily activities* of people. Daily activities vary for each individual and they have an impact on responding to a delivered intervetion and performing the actual behavior. A concept called *activity timeline* is introduced to represent all the daily activities of a person during the day from wake up to sleep. Using the activity timelines, activities for each simulated person are generated in a randomized way. The third dimension is *reactions to the delivered interventions*, which is affected by the suitability of daily activities concerning the engagement with interventions as well as preferences of people on specific interventions types. The last one is the actual *behavior performance*. It is determined by the prediction on remembering the behavior by the habit formation model and suitability of the daily activity for performing the behavior.

The aforementioned four distinguishing factors are configured for four personas via several rules and conditions. The learning algorithm is expected to adapt the intervention delivery strategy in a way that captures the varying configurations of these simulation parameters. Details about the persona-specific configurations of the simulation parameters are presented in the subsequent sections.

7.2.3.1 Configuring the Habit Formation Model

The habit formation model had been summarized in the preliminaries section stating that the model needs two external inputs for its initialization, which are *be-*

havior frequency and *commitment intensity*. Considering that both parameters can take values between 0 and 1, the following values are set to $0.2, 0.4, 0.6$ and 0.8 for $Person-1, Person-2, Person-3$ and $Person-4$ respectively. Having a relatively higher commitment intensity is an indicator of giving more importance to or having more desire for the targeted behavior. From the opposite perspective, having a relatively less commitment intensity might indicate that a person would be unwilling to perform the behavior without expecting long-term benefits. It might also be the case that the person might be finding the behavior difficult or demanding [102, 103]. For simplicity, the initial behavior frequencies are initiated with the corresponding person's commitment intensity value.

7.2.3.2 Simulating Daily Activities

Daily activities are simulated via activity timelines. The aim of having such timelines is to simulate states that are (or not) suitable for engaging with interventions and to simulate states that are suitable (or not) for performing the activities described in action plans. A timeline is a placeholder for a sequence of activities to be performed throughout the day. Timelines are populated with predefined activities, with a fixed order, that can be semi-randomized for each person for each simulated day. The semi-randomization is realized in the following way: Firstly, a subset of the activities are selected among the initial complete set activities. Then, each activity is associated with a *start time* and *duration*. The sequence of activities differ for each learning episode, i.e. a simulated day, and from person to person. Each activity contains information regarding the *time, location, physical activity, emotional status* and *phone screen status* of the person. An example activity is presented in Table 7.1, containing possible values for the activity parameters (middle column) and specific instantiations of the parameters (right column). Emotional status is selected from a predefined set of options with a certain probability. According to the example, the person would have neutral emotional status with a probability of 70%. Most of the time, an activity does not have a fixed start time, but has a relative one depending on the duration of the previous activity. The duration is also not fixed. In the example, the duration of the activity can be between 30 minutes and 60 minutes.

Table 7.1: Elements of a daily activity

Daily Acitivity Element	**Possible Values**	**Example Value**
activity description	<Verbal description of the activity>	checking daily news and e-mails before start working
location	Home / Office / Outside	OFFICE
physical activity	Sedentary / Walking / Running / Indoor_Activity / Driving	Sedentary
phone usage	Active / Screen_Off	Active
emotional status	Emotional status: Neutral / Relaxed / Angry / Stressed	Neutral = 70% Stressed = 15% Angry= 15%
start time	A *relative* time or fixed time point	Relative
start time variation	A contextually (considering the daily activities) appropriate number	0
duration	A contextually (considering the daily activities) appropriate number	45 minutes
duration variation	A contextually (considering the daily activities) appropriate number	15 minutes
behavior performance suitability	Yes / No	Yes

Table 7.2: Preferences associated to simulated personas on intervention types

	Intervention-1	Intervention-2	Intervention-3
Person 1	0%	50%	10%
Person 2	20%	80%	80%
Person 3	80%	10%	80%
Person 4	50%	50%	50%

7.2.3.3 Simulating Reactions to Interventions

Personal preferences on specific intervention types and suitability of the daily activity are the determinants of engaging with an intervention. Personal preferences on intervention types are simply represented with a percentage that is used as the probability of engaging with an intervention. Persona-specific preferences on intervention types are given in Table 7.2. The percentages are basically a reflection of personal traits on appreciation of the interventions by the person. For example, an intervention content may include social interaction with other people, which may not be preferred by the person. The values have been selected heuristically to be able to observe contrasting values based on differing preferences of the personas on the same intervention types.

These probabilities are considered only if the current daily activity is also suitable for checking the phone and engaging with the intervention. In this respect, three of the daily activity parameters namely *emotional status, physical activity status* and *phone screen status* are used as determinants for engaging with an intervention. Specifically, people are assumed to engage with an intervention when they are sedentary, have a neutral / relaxed emotional mood and when their phone screen is on.

7.2.3.4 Simulating Behavior Performance

Whether the person would perform the behavior or not is actually the outcome of the habit formation model. If it is positive, it is assumed that the person would perform the behavior during the activity. Some of the activities included in the activity timelines are marked as as suitable for performing the behavior. For this purpose, the *behavior_performance_suitability* parameter is set accordingly as seen in Table 7.1.

7.2.4 Hypotheses

Driven by the benchmarked parameters of simulated personas, the following hypotheses are introduced as conditions for which the JITAI personalization algorithm is expected to adapt intervention delivery accordingly. For each hypothesis, the corresponding simulation parameters leading to the hypothesis are also presented:

- ***Varying parameter:*** According to the habit formation model referred in the current study, people with higher commitment intensities perform the behavior more frequently and reaches maximum habit strength faster.
  Hypothesis 1: *The proposed approach should deliver interventions throughout a longer period of time as the commitment intensity associated with a persona decreases.*
- ***Varying parameter:*** A well-formed habit (i.e. automatic performance of the new behavior) indicates that the person performs the behavior with less dependence on extrinsic reminders and motivators.
  Hypothesis 2: *The number of delivered interventions should be inversely proportional with the perceived habit strength and become more and more intermittent throughout the simulated care process.*
- ***Varying parameter:*** As a reflection of their individual differences and preferences, the simulated people favor different intervention types.
  Hypothesis 3: *The ratio of selected interventions should be proportional to the preferences of the persons.*
- ***Varying parameter:*** Daily activities of each person are generated semi-randomly based on personal activity timelines. Therefore, in addition to that each person has distinct activities during the day, the activities for the same person vary among the simulated days because of the randomness included in the activity generation mechanism. Varying daily activities determine both when the behavior could be performed and when the intervention (i.e. the mobile phone

notification) can be engaged with.

Hypothesis 4: *The proposed approach should deliver interventions respecting to people's daily life patterns by respecting the timing and suitability of activities for engaging with the intervention and performing the behavior.*

7.3 Results

In the light of the experimental setup presented in the previous section, the results are presented from two main perspectives. First, concerning the suitability of the proposed approach for personalization of JITAIs, how the experiment results validate the proposed hypotheses is discussed. It is claimed that the proposed approach adapts intervention delivery strategy in terms of the *adaptivity (i.e. type and frequency of interventions)* and *just-in-timeness (i.e. timing of interventions)* aspects. Second, raw results from the simulated experiments are presented intending to show improvements beyond the base RL algorithm and the *jump-start* challenge being addressed by the proposed approach. The results are analyzed with respect to the evaluation metrics commonly used in the transfer learning domain.

Eight experiments are conducted such that each experiment contains 100 trials and each trial contains 100 learning episodes. The first four experiments are configured for Person-1, Person-2, Person-3 and Person-4 respectively. They do not utilize the *State Classifier* as it has not been trained yet. The aggregated data throughout the first four experiments are used to train the State Classifier. The fifth to eighth experiments are conducted again for Person-1, Person-2, Person-3 and Person-4 respectively, but now utilizing also the State Classifier. All the results presented below are average results obtained from the experiments 5-8. The results of the individual experiments are used in cases where person-specific results are presented. Otherwise, the data of the experiments are merged, and the results are compiled out of the merged data.

7.3.1 Validation of Hypotheses

Hypothesis-1 validation: The commitment intensities are set in an increasing manner from Person-1 to Person-4. This implies that the time required to form habit for the targeted behavior should be the shortest for Person-4 and the longest for Person-1. The left-y versus x-axis of Fig. 7.1 represents the number of interventions delivered in each episode. The dappled lines with square shapes show the values obtained for this metric. As hypothesized the number of interventions declines to vicinity of zero the latest for Person-4, i.e. the algorithm delivers interventions for a longer time for Person-4 compared to other persons. The length of the intervention delivery period decreases for each person inversily proportional to the commitment intensity values set for the associated personas.

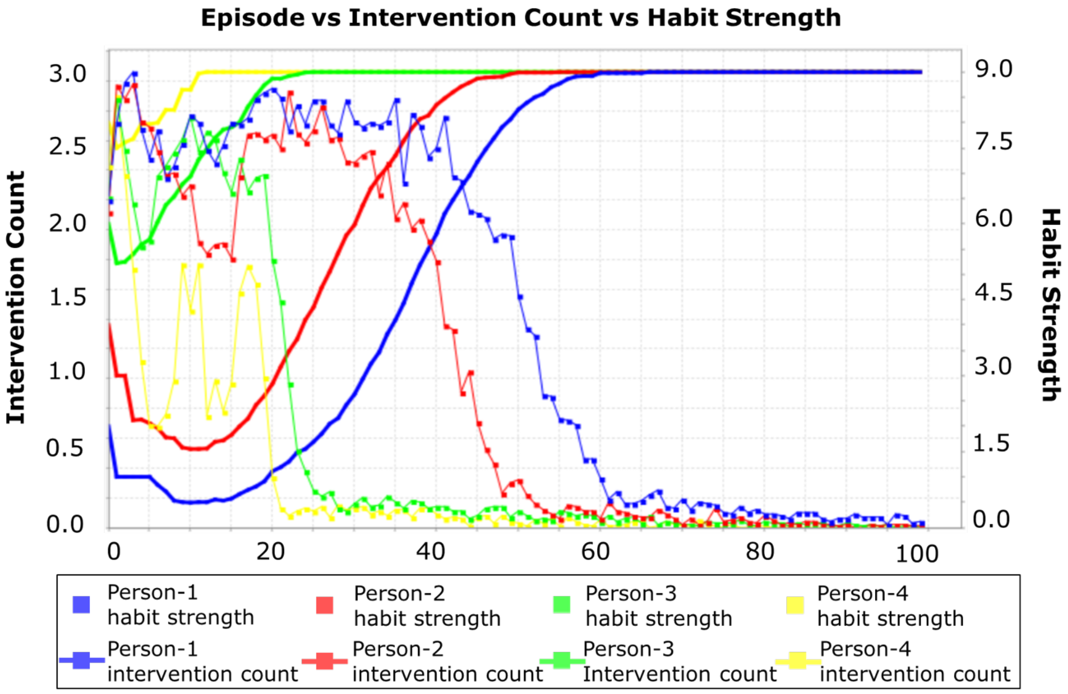

Figure 7.1: Episode vs intervention count vs habit strength plot

Right-y axis versus x-axis of Fig. 7.1 represents the simulated habit strength per episode. The time elapsed to reach the maximum habit strength for each person is also in consistency with the commitment intensities set previously such that Person-1 reaches the latest and Person-4 reaches the earliest. Considering the length of the duration to reach the maximum habit strength, the values obtained in our study are congruent with the results obtained by Lally et al. [83]. They develop an habit formation model on empirical data regarding the performance of targeted behavior and habit strength perceived by the subjects. Their model outputs the duration for forming habit on the targeted behaviors where behaviours are in varying complexities. According to Lally et al., the duration for habit formation varies from almost a month to a few months where the complexity of behavior is the main determinant. In this study, only one behavior type is considered. However, as stated earlier, studies show that the same behavior might be perceived in varying difficulties by different people. Considering all these factors, it is argued that the simulation is realistic in terms of generation of habit strength values i.e. performing the targeted behavior without needing extrinsic reminders.

Hypothesis-2 validation: Looking again at the dappled lines of Fig. 7.1 showing the number of delivered interventions, the averages are relatively high in the starting phase of the learning process for each persona. Although there are fluctuations, during this initial phase, it can be observed that the average numbers of delivered interventions for Person-1, the dapped blue line, are higher than the others most of the time. That is, Person-1 who has the lowest commitment intensity receives the highest number of interventions. Furthermore, as shown in Fig. 7.2-a, the total number of interventions delivered for Person-1 is higher than the rest. After a while in Fig. 7.1, the number of delivered intervention decreases to vicinity of zero. As envisioned by the second hypothesis, the number of delivered interventions reacts to the changing habit strength. The number of delivered interventions is inversely proportional with the perceived habit strength. It can also be observed that the change rate of intervention throughput of the algorithm is inversely correlated with the change rate of the simulated habit strength. That is, as the increase rate of habit strength increases, decrease rate of the number of intervension increases in similar amounts.

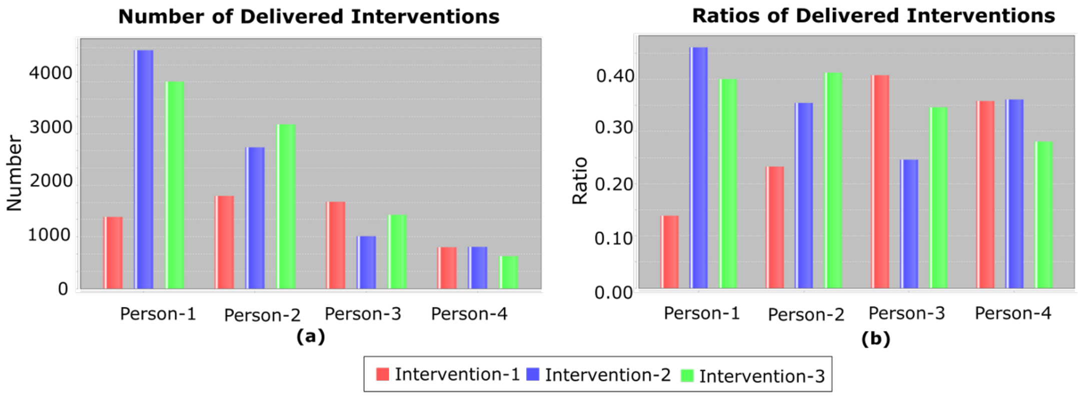

Figure 7.2: Person vs intervention type ratio plot

Hypothesis-3 validation: Before validating the obtained results against the simulation configurations and hypotheses, it should be reminded that the three simulated intervention types are clustered into the two disjoint sets according to their categories. That is, only the interventions of the same set can be alternatives to each other. Therefore, the results should be evaluated independently for each set. Being reminder interventions, Intervention-1 and Intervention-2 belongs to the *reminder* set, whereas Intervention-3 belongs the *motivation* set.

To be able to validate the reflection of preferences of different persons on the same intervention, the intervention ratios i.e. Fig. 7.2-b should be analyzed. The ratios obtained for Intervention-1 reflect the specific preferences i.e. 0%, 20%, 80% and 50% for Person-1 to Person-4 respectively. However, Intervention-2 and Intervention-3 do not reflect the initial preferences. This is because, actually, the preferences of an individual person on intervention types do not sum up to 1. As previously stated earlier, each specific preference value is used as a probability when the user is presented with the associated intervention type. For example, Person-1 never reacts to Intervention-1 or Person-2 reacts to Intervention-2 or Intervention-3 80% of the time upon encountering with one of these interventions. What should be evaluated is the comparison of intervention ratios among each distinct set per individual basis. From this point of view, ratios of Intervention-1 and Intervention-2 reflect the initial prefences for all persons. For example, the ratio of Intervention-1 ratio is higher than the ratio of Intervention-2 for Person-3, considering that preferences of Person-3 is 80%

and 10% for Intervention-1 and Intervention-2 respectively; ratio of Intervention-1 is higher than the ratio of Intervention-2 for Person-3; ratios of Intervention-1 and Intervention-2 are almost equal to each other.

At the first glance, though, it can be said that there are inconsistencies between preferences and the obtained results. For example, Person-1 has no interest at all for Intervention-1 and little interest for Intervention-3. But, the results show that the ratios for Intervention-1 and Intervention-3 are 15% and 40% respectively. In case of Intervention-1, the main reason of the inconsistency is random selection of interventions in unknown states i.e. the states that have not encountered before by the learning agent. That is, the learning agent make random decisions on selecting Intervention-1 or Intervention-2. This problem applies also to Intervention-3, which is the only alternative in the relevant intervention set. In addition to that, the algorithm favors intervention delivery regardless of its type, in case the person is predicted not to perform the behavior. In such cases, any type of intervention is considered as a cue remanding the behavior.

Hypothesis-4 validation: According to the design of the simulation, intervention delivery and behavior performance do not necessarily happen at the same time. Most of the time there is a temporal difference between the two. Fig. 7.3 shows the ratio of interventions that fall into pre-defined ranges of such temporal differences. An individual range is represented at with a bar in the plot. Overall, the figure validates the hypothesis such that the ratio of delivered interventions decreases as the time difference increases except the 31-60 minute bar. For 74% of the interventions, the time difference is not more than 30 minutes. The violation of the pattern arises from the daily activties that are suitable for intervention delivery but not behavior performance. This is actually an indicator of the algorithm's capability on learning the personalized patterns.

Furthormore, the results also capture the conditions on persons' contextual parameters set for the intervention engagement. Specifically, interventions have been sent in suitable conditions such that in physically sedentary mode; in a convenient emotional mode and when the phone screen is on 76%, 80% and 69% of the time respectively.

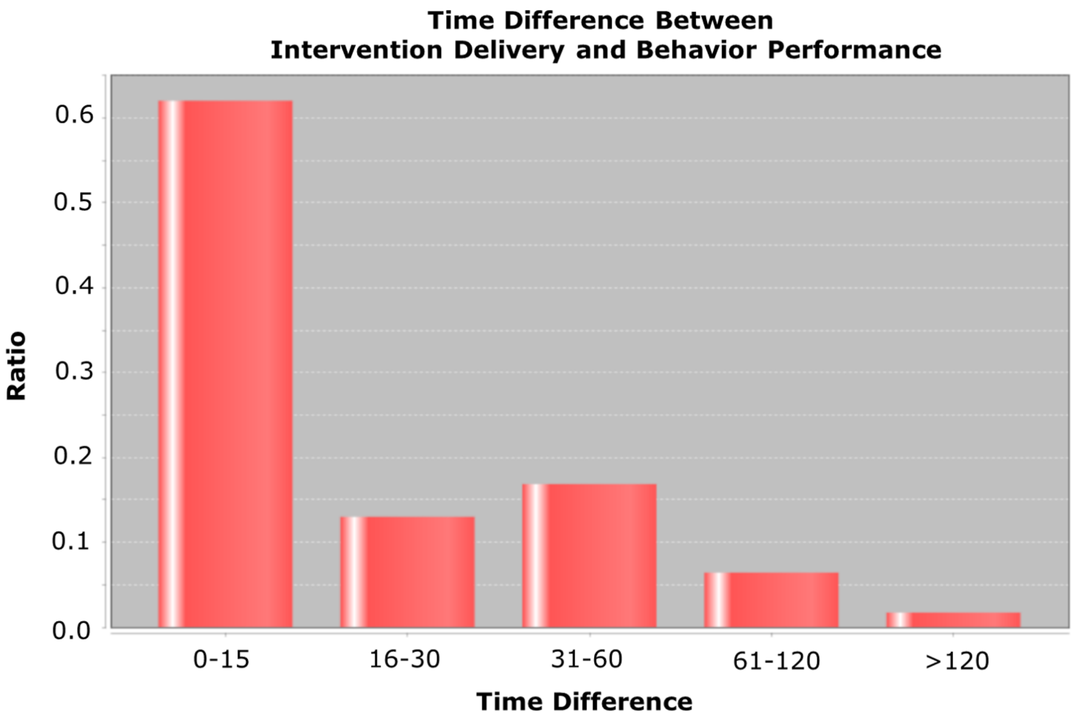

Figure 7.3: Difference between JITAI delivery and behavior performance times

7.3.2 Improvements on the Base Algorithm

In this section, results both for the intervention-selection and opportune-moment-identification models are presented. Concerning the former model, results obtained with the base RL algorithm, i.e. Q-Learning, are presented. Based on the results, it is argued that the base algorithms converge sufficiently fast considering the relatively low complexity of the model. Then, comparative results obtained by the opportune-moment-identification model are presented. The base RL algorithm is compared with the two extended versions using the modified eligibility traces and transfer learning methods. It is argued that the extended versions yield better results than the base algorithm.

7.3.2.1 Results for the Intervention-Selection Model

Fig. 7.4 shows the raw rewards aggregated per episode per person. The plot shows that the rewards correlate with the habit strength plot as depicted in Fig. 7.1, indicat-

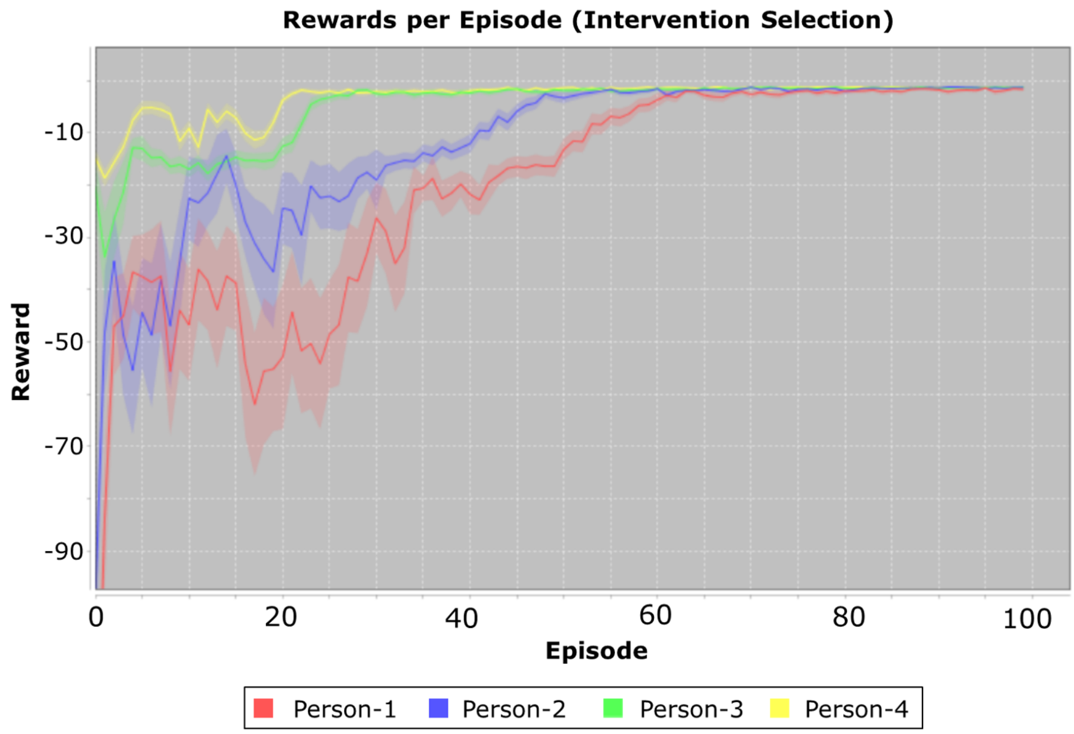

Figure 7.4: Rewards per episode in the intervention selection model

ing that the perceived habit strength is a strong determinant of the obtained rewards. This is an expected effect as the algorithm is expected to make the intervention delivery more intermittent as the perceived habit strength increases.

As can be seen in the figure the total reward amount reaches a plateau before the trial ends. This is an acceptable convergence time considering the targeted specific domain. For example, for diabetes, although specific studies claim that two-weekly doctor visits yield faster achievement of targeted clinical outcomes [104], standard-based guidelines recommend 3-monthly or 6-monthly visits [98]. Therefore, it is argued that achieving a convergence in approximately 100 episodes is satisfactory in the scope of this study.

Intervention selection model deals with adaptation of intervention delivery with respect to frequency and type of interventions. Concerning the frequency aspect, it has been already shown that the number of delivered interventions is adjusted based on people's perceived habit strength in Fig. 7.1. Fig. 7.2-a also shows that the personal preferences have been reflected on the selected intervention types. Complementary to these results, Fig. 7.5 shows the reaction ratio, representing the ratio of number of interventions engaged with to the total number of delivered interventions per episode per person. The results show that the reaction ratio is around 50% in average for all personas during periods where interventions are actively delivered. The reaction ratios point out potential improvements for the intervention-selection model. In the discussion section, some potential improvements are discussed to obtain better results.

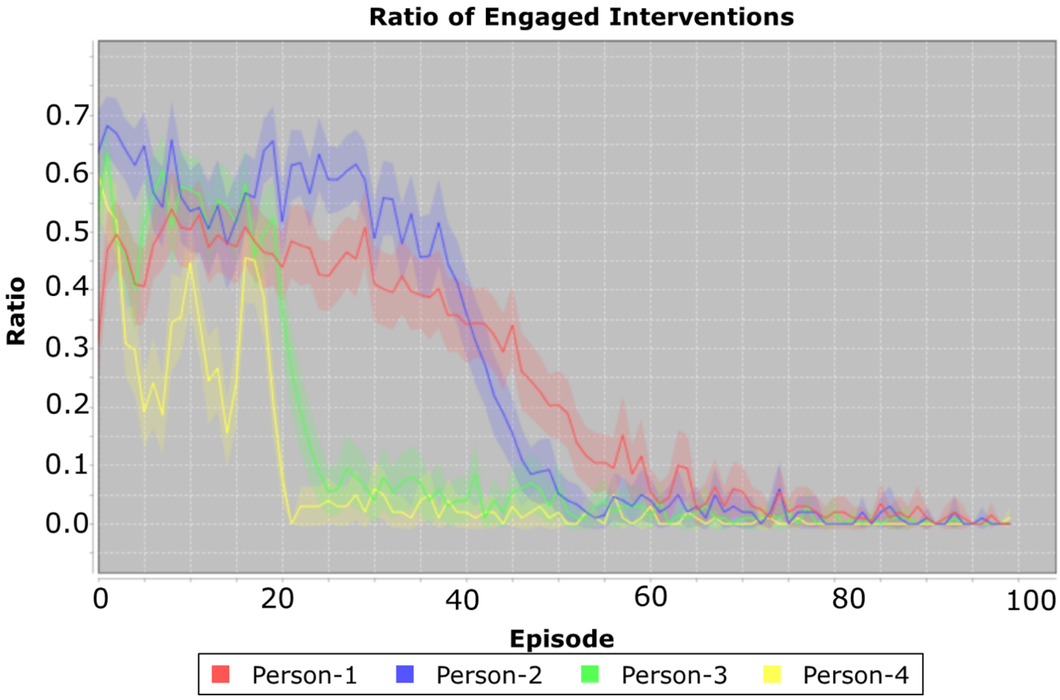

Figure 7.5: Ratio of engaged interventions

Complexity of the Model: These results have been obtained using the standard Q-Learning algorithm. As an implication of the complexity of the model, it should be noted that the total number of distinct states visited by the learning agent is 96 in average among the trials. As will be seen in the next section, this number is much higher for the opportune-moment-identification model, which necessitates improvements on the base algorithms.

7.3.2.2 Results for the Opportune-Moment-Identification Model

In this section, results obtained for the opportune-moment-identification model are discussed with respect to the performance metrics introduced by the transfer learning methodology, namely: *jump-start, asymptotic performance, total reward, transfer ratio* and *time-to-threshold*.

Fig. 7.6 addresses all these metrics simultaneously. The figure shows the rewards aggregated per episode. Rewards are shown for 3 versions of the learning algorithm utilized by the opportune-moment-identification model. The green line shows the results for the base Q-Learning *(QL)* algorithm. The blue line shows the results for the base algorithm improved with the selective eligibility traces *(QL-SET)*. Lastly, the red line shows the results for the base algorithm improved with both the selective eligibility traces and transfer learning mechanism *(QL-SET-TL)*. Furthermore, the yellow line shows the rewards obtained via the fixed delivery schedule.

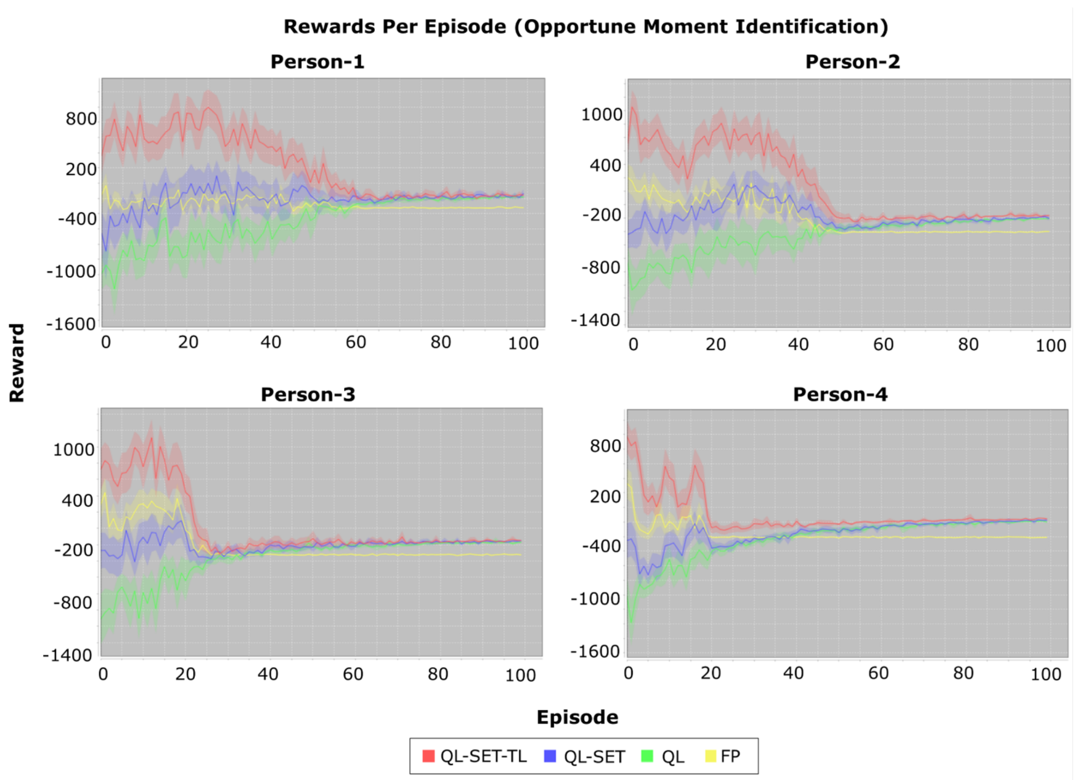

Figure 7.6: Rewards collected per episode in the opportune-moment-identification model

The blue line laying over the green one, indicates a better jump-start achieved by the selective eligibility traces compared to the base algorithm. Nevertheless, the red line outperforms the other two with a notable highest jump start at the beginning of the learning process. The area under the lines correspond to the total rewards collected by each algorithm. According to the figure, QL-SET-TL collects more reward than the others throughout the experiment. This also implies that QL-SET-TL is more effective than the others considering the ratio of number of engaged interventions to the total of number interventions sent. With respect to the asymptotic performance, QL-SET-TL has a higher course than the other two during the active intervention delivery periods. Concerning the time-to-threshold metric, no specific threshold is defined. However, it can be said that all versions stabilize almost at the same time. Lastly, the area between the red line (QL-SET-TL) and blue line (QL-SET) correspond to the rewards that were collected thanks to the transfer learning. In this respect, considering the active intervention delivery period, the area between the blue and red lines indicate a notable contribution of transfer learning on the collected rewards. The results also show that the fixed delivery schedule performed the worst compared to all the versions of the RL-based algorithm.

The results presented in Fig. 7.6 shows that QL-SET-TL performs the best at the beginning of the learning process. In this respect, Fig. 7.7 shows the ratio of actions determined by *State Classifier* versus the agent's internal policy. The State Classifier selects most of the actions at the beginning of trials. The number of actions selected by the agent's internal policy increases gradually as the agent itself learns more and more about its environment.

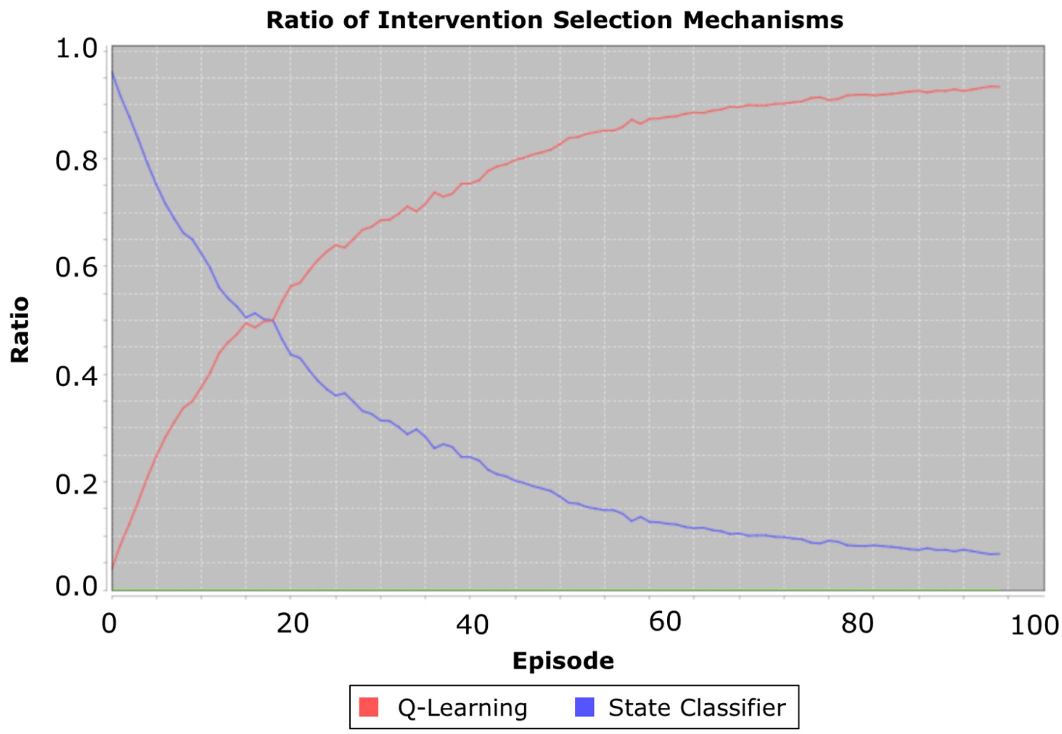

Figure 7.7: Ratio of actions per action selection mechanism

Complexity of the Model: The number of distinct states that the agent visits in this model is 2596 in average, indicating a quite larger state set compared to the intervention-selection model. Although the complexity of the model has increased, it still performs well considering the timing differences between the intervention delivery and behavior performance times as depicted in Fig. 7.3. Also, despite the higher complexity of the model, QL-SET-TL obtains relatively higher rewards. These results substantiate the role of transfer learning on selecting better actions.

7.3.3 Real-Life Case Study

In addition to the simulated case study, a small-scale real-world experiment has also been conducted, where a preliminary version of the opportune-moment-identification algorithm was in place to break the sedentary behaviors of office workers during the working hours [91]. In that study, though, SARSA[105] as the base RL algorithm was used instead of Q-Learning.

Table 7.3: Real-world experiment results

	Control-fix	*Control-SARSA*	*Focus*
Average reaction time in seconds	2252	500	712
Total number of engaged interventions	10	67	72
Ratio of engaged interventions to the total number of interventions	3%	23%	38%
Average number of daily interventions (per user)	4	7	2

17 office workers have been recruited to make them more physically active during their work life. Specifically, the aim was to motivate them to take a break via interventions in the form of mobile app notifications. The mobile app collects their contextual data (wi-fi, activity and phone screen status) during the day. Upon an update on one of these parameters, the algorithm decides in real time to deliver an intervention or not. It updates its internal policy by processing the user reactions that could be: discarding an intervention, seeing the intervention details or clicking one of the two buttons (positive / negative) located in the intervention detail page indicating their willingness to perform the suggested activity.

The participants are divided into three groups namely; *control-fix, control-sarsa* and *focus*. Control-fix group receives interventions 4 times a day such that at least one hour after starting to work (morning and afternoon) and at least one hour after the previous intervention. Control-SARSA and focus groups receive interventions dynamically based on the decisions of SARSA and SARSA-SET-TL respectively. Results obtained from an almost 2 weeks of experiment are presented in Table 7.3.

As expected, even for this small, initial experiment significantly better results have been obtained by the proposed approach. Both, SARSA and SARSA-SET-TL yield better results than the fixed schedule considering the number of engaged interventions and their ratio to the total number of interventions sent. Reaction times for the fixed schedule is also quite high compared to the dynamic delivery mechanisms indicating that dynamic algorithms were better in adjusting the timing of interventions. Comparing the base SARSA and SARSA-SET-TL, the latter outperforms on the ratio of engaged interventions, validating the proposed improvements on the base algorithms. The reaction time is higher for SARSA-SET-TL, as if it performed worse in terms of timing of interventions. However, the SARSA algorithm sends much more interventions than SARSA-SET-TL, that are probably seen by users earlier. This assumption is consistent with the lower engagement ratios of SARSA.

7.3.4 Validation of the Reward Function Instantiations

Reward functions of the RL models are critical as they guide the learning agent in the correct direction by generating appropriate amount of rewards considering the appropriateness of the action taken. Therefore, instantiation of the reward values are critical in terms of representation of the desirability of actions. Reward constants should be set considering the *Learning rate* parameter of RL systems, which is used to adjust the impact of recently observed rewards on the accumulated knowledge (Eq. 3.2). As the learning rate increases, the impact of the recent reward also increases and it affects the internal policy more. According to the way the learning agents are supposed to behave, some conditions in the reward functions are desired and some of them are not. While specifying reward values for each condition zero (0) has been accepted as the neutral point and the magnitude of the rewards have been determined according to the desirability of the conditions. However, although specific values have picked as reward values, they are not definitive and they can be changed by shifting the neutral point and scaling the values as long as the relations among the reward values are preserved in terms of the magnitude and sign of rewards. In the two subsequent sections, the details about this approach are presented along with some experimental results obtained with varying reward values.

7.3.5 Validation of the Opportune-Moment-Identification Model Reward Function

The reward function of the opportune-moment-identification model includes two equations (Eq. 5.1 and Eq.5.2), both of which include 3 constants. Eq. 5.1 is conditioned on two variables namely the action taken (omi_a_t) and reaction to the delivered intervention ($reacted_t$). Its main aim is to produce a reward indicating whether the state where an intervention is delivered is an appropriate state for engaging with the intervention or not. Thus, considering the variables included, the case where an intervention is delivered as an action and the person's reaction to the intervention is positive i.e. ($omi\_a_t = Deliver\_Interventionandreaction_t = true$) is desired. Therefore, in case an intervention is delivered but the reaction is negative i.e. ($omi\_a_t = Deliver\_Interventionandreaction_t = true$), a relatively smaller reward should be generated. The relative difference between the rewards generated for these two cases has the following implications. From a domain-specific perspective, it shows the importance given to the two cases. In this manner, discovering a state that is appropriate for engaging with an intervention is more critical. However, there might be cases that are appropriate for engagement but no engagement occurs because of the randomization factors. Such cases might be observed successively and in this situation the burden created on the person is represented by multiplying negative reward with the $number\_of\_attempts_t$ variable. If the sequence of such cases gets longer, the magnitude of the negative reward increases. Thus, to neutralize or even revert the incorrect knowledge learnt in such situations quickly, $sent\_reacted$ value has chosen relatively higher (1000). Nevertheless, it could have been chosen in different amounts resulting with a similar performance as long as it would be the largest value among the rewards specified for the cases of the equation. Fig. 7.8, specifically (a), (b), (c) and (d) parts, show the rewards obtained for Person-1 for the values 50, 500, 2500 and 10000 respectively while the other reward values are fixed. Although the results obtained with 500 and 2500 values provide similar results in terms of learning performances of the benchmarked algorithms. The pattern changes towards to the edge cases. In case an even larger reward is specified, the learning performances of QL-SET and QL-SET-TL overlap as the effect of negative cases become negligible. On the other side, setting a smaller value for $sent\_reacted$ weakens its effect. As a result, since the number of undesired actions taken by QL-SET is fewer compared to QL-SET-TL, performance plots of these two algorithms diverge and the amount of rewards obtained by QL-SET get closer to the base QL. Modifying only the $sent\_not\_reacted$ reward creates just the opposite effect created by $sent\_reacted$.

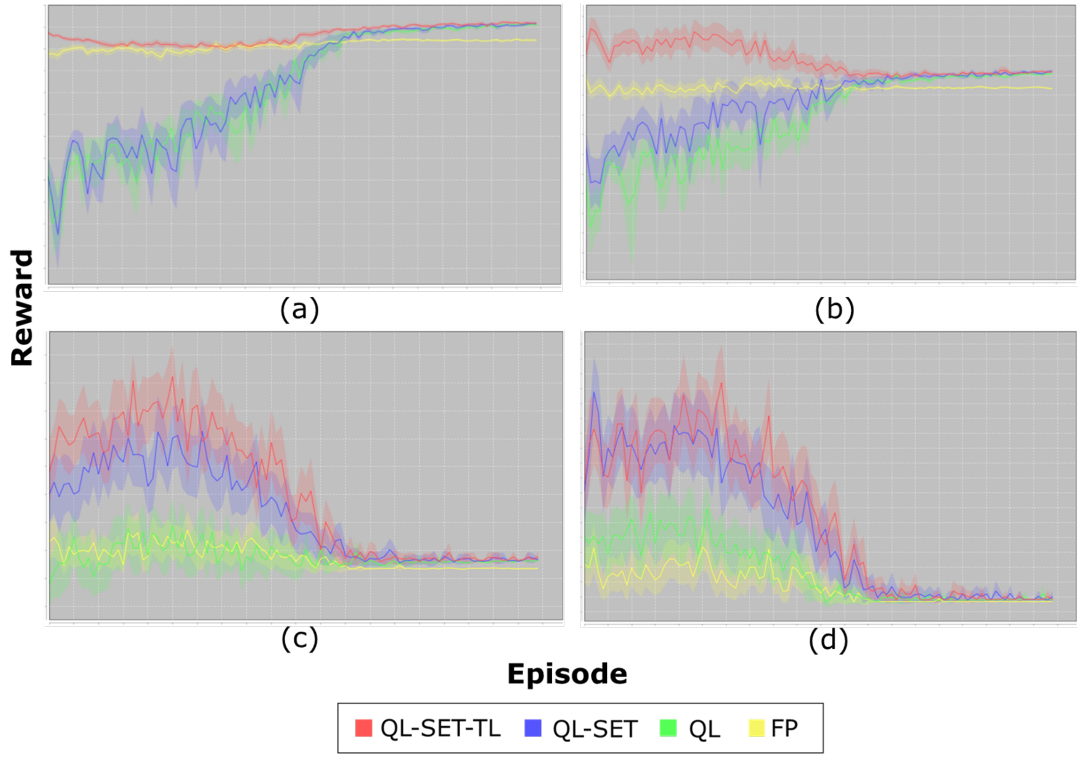

Figure 7.8: Rewards obtained with the modified the sent-reacted variable in Eq. 5.1

The intuitive choice for the not_sent reward specified for the cases with no intervention delivery would be 0 since there is no feedback available from the person. However, in this case, the learning agent gets stuck at a local minima by always selecting the $Deliver\_Nothing$ action. In order to prevent this scenario, a nominal negative reward is generated to force the agent take the $Deliver\_Intervention$ after some time.

The second equation generates a value to fine-tune reward generated for the desired case. Basically, the temporal difference between the intervention delivery and behavior performance is considered. The aim is to have this difference as smallest as possible to increase the effectiveness of the intervention. Therefore, the values are picked in a decreasing manner as the difference gets higher. Providing that this pattern is preserved while specifying reward values in to scope of this equation, the results are also maintained. Fig. 7.9, specifically (a), (b), (c) and (d) parts, show the results obtained when the reward values are multiplied by 10, 25, 50 and 125. As can be seen from the figure, the results obtained for the first three scales reflect the initial patterns, even though the lines get closer as the scaling increases. However, increasing the scale even further causes all the results to overlap. This is because the rewards generated in the first equation become negligible.

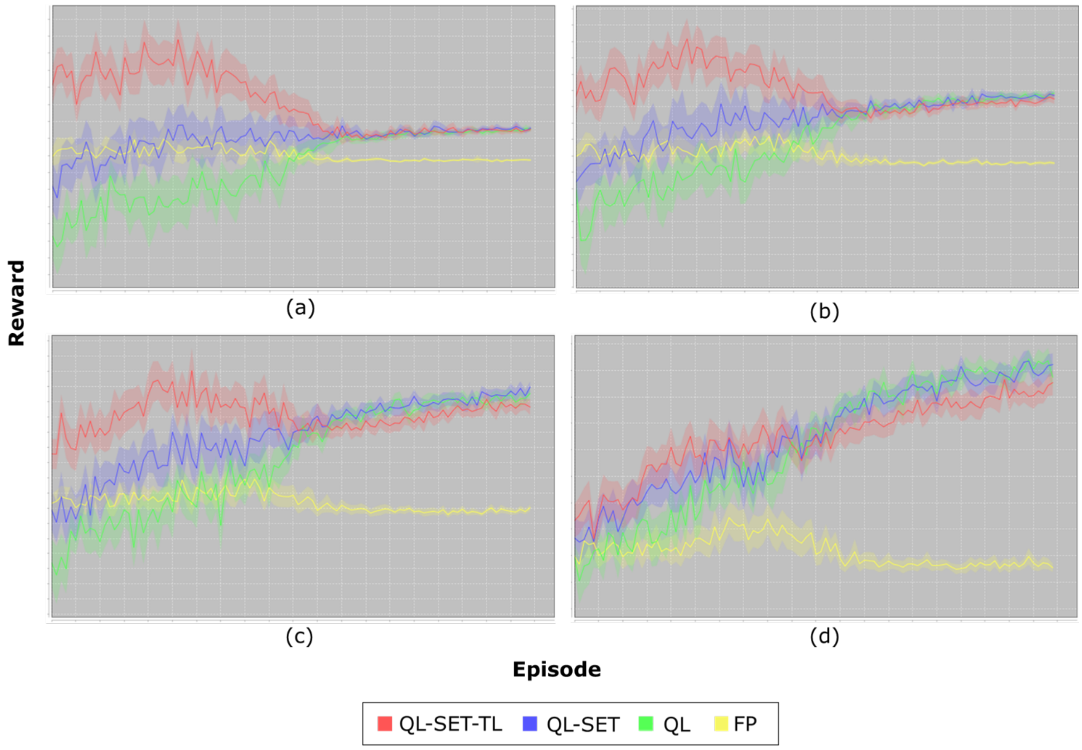

Figure 7.9: Rewards obtained with varying scales of temporal rewards

7.3.6 Validation of the Intervention-Selection Model Reward Function

Similarly, the reward function of the intervention-selection model is also a conditional equation (Eq. 5.3) depending on 3 variables namely delivered action (is_a_t), remembering the behavior ($remember\_behavior_t$) and reaction to the delivered intervention ($reacted_t$). The reward values for the conditions of this equation have basically identified based on the intervention selection policy that we want the learning agent to come up with.

Considering the habit formation model, situational cues should be provided to people if they forget to perform the behavior. Thus, the cases where the person does not remember to perform the behavior i.e. ($remember\_behavior = false$) is critical. Among the cases satisfying this main condition, the case where no intervention is delivered is the most critical one because it leads to non-performance of the behavior at all. Therefore, a relatively large, negative reward is specified for this case (-50). There are two more cases where the person does not remember to perform the behavior. These two remaining cases include intervention delivery ($is\_a_t \neq No\_Intervention$). However, engagement may or may not occur. The case where an engagement occurs is a desired case so, a positive reward (10) is set for this case. Since the non-engagement cases are considered as burden on the person, they are negatively rewarded. However, since the person forgets to perform the behavior, the reward specified for this case (-5) is larger than the reward specified for the case where the person remembers to perform the behavior and does not engage with the intervention (-10).

In the first case of the equation, the person remembers the behavior and also reacts to the delivered intervention ($remember\_behavior_t = true\ and\ reacted_t = true$). However, this is still accepted as a burden on the person since s/he does not need to be reminded. Therefore, the reward is specified as 3 for this case.

Lastly, -1 reward is specified for the case where the person remembers the behavior and no intervention is delivered. Similar to the case of the opportune-moment-identification mode, there is no feedback available. Therefore, although 0 reward for this case seems intuitive, in order to prevent getting stuck in a local minima, 1 reward is specified for this case.

Scaling the reward values all together does not change the algorithm results. Furthermore, similar to the opportune-moment-identification model, the overall reward patterns are obtained with individually modified reward values as long as the magnitude and sign conditions are preserved.

CHAPTER 8

DISCUSSION

Legitimacy of the approach: The dynamic JITAI personalization mechanism and multi-dimensional expandability of the JITAI design mechanism are the two innovative characteristics of the proposed approach advancing the state-of-the-art research. Both characteristics bring opportunities for adoption of the system by various entities such as mobile/web application vendors in the behavioral health software market, public health organizations or other healthcare organizations working on clinical studies aiming at large scale digital interventions in patient populations.

From a perspective of improved care programs, the proposed approach has a critical importance. Patients can strive for their health by themselves. In this respect, behavioral lifestyle patterns are important predictors of health outcomes such that patients can reduce the risk of chronic diseases by adopting healthier lifestyles. The evidence is overwhelming that physical activity and diet can reduce the risk of developing numerous chronic diseases and in many cases even reverse the existing disease [7]. As introduced earlier, personalized support is a critical enabler of increasing the adherence to behaviors like physical activity or diet. Therefore, the proposed algorithm with the capability of adapting itself towards a personalized intervention delivery strategy is a valuable tool for behavior change programs.

Capturing the rules associated with the simulated concepts: Simulation results have already been discussed in the previous section by describing how they are aligned with the hypotheses and how they deviate from the expected results along with the causes of deviations. As a summary, it is argued that the proposed algorithm is able to capture the rules that are associated with the simulated concepts.

Further improvements: Despite the innovative character of the current approach, there is room for potential improvements. For example, as mentioned earlier, the fluctuations of the intervention counts in Fig. 7.1 happen when the learning algorithm encounters with unknown states, i.e. the cold-start problem. Tackling with the cold-start problem, the evidence formed in various micro-randomized trials[106], measuring the effect of individual intervention components, or any expert knowledge-based heuristic can be utilized to provide a warm-start for the learning algorithm. An unknown state might be encountered for example when the person would reach to a certain habit strength, the highest one, for the first time. Instead of taking random actions in such cases, the algorithm might employ a machine learning classifier to make an educated guess or it may simply favor not delivering an intervention in proportion with the current habit strength. The RL methodology is convenient for integration of such external knowledge by setting the initial scores of relevant state-action pairs inside the learning agent's policy accordingly.

In the scope of this study, only an initial basis has been established concerning the model parameters that are utilized to capture the personalized patterns in relative contexts. Therefore, it is not claimed that those are the necessary/sufficient set of parameters. On the contrary, the models could be enriched with additional contextual parameters pertaining to environment, mobile phone or person him/herself to represent the person more accurately. Such an improvement may in turn require further optimization of the algorithm, e.g. identification of covariant parameters, for its applicability on the personalized care domain concerning the performance.

Specific to the intervention-selection model, as described earlier, the intervention-selection model is mainly based on the mathematical habit formation model. Although, currently a model-free approach has been implemented, the model can be designed as a model-based system. This enables training of the value-function of the RL environment via intermediate simulations before taking an action. Even, the intervention-selection model could be split into two separate models such that the first one would learn on delivering an intervention or not for a specific activity prescribed in the action plan. Note that this differs from the opportune-moment-identification model, which consider momentary parameters to deliver the selected intervention or not. The second model would just learn the preferences of people on the intervention types. This might lead to better results considering the selection of interventions types.

Besides the improvements on the learning models, the JITAI design capabilities could be improved with additional intervention types or content presentation modalities. Appendix A presents all the currently available constructs of the design approach. Albeit being a simple system concerning the limited number of built-in constructs targeting the POWER2DM case study, the design mechanism lays out the basis to expand the system with more constructs as needed by the targeted health problem.

A limitation of the study is the limited scope of the simulation. A limited number parameters are considered as differentiating factors in the persona simulation including preferences on intervention types, commitment intensities and daily activities. However, a more realistic simulation could be achieved by also considering factors like self-efficacy, motivation, prior experience changing the behavior or outcome expectancies.

CHAPTER 9

CONCLUSION

The main outcome of this study is a framework that can be utilized for JITAI design and personalization. The framework can be customized for specific care programs targeting varying health problems and populations. The design mechanism, incorporating a rule definition language, can be specialized with add-on constructs to conceive interventions addressing the specific requirements of a care program. The personalization part employs a reinforcement learning based approach to optimize/personalize the intervention delivery concerning the frequency, type and timing of interventions dynamically according to the data aggregated for people over time.

The JITAI design mechanism has been validated by providing example JITAI definitions where the characteristics of JITAIs are extracted from various relevant resources available in the literature such as clinical guidelines and taxonomies of behavior change; and from a real-world case study providing self-management support to diabetes patients. The personalized intervention delivery mechanism has been validated through simulated and real-life case studies. Although the real-life case study has been performed with a preliminary version of the personalization algorithm, it has yielded better results compared to a fixed intervention delivery strategy. In the scope of simulated case study; action plans, JITAIs and personas, with differentiating characteristics are simulated. The obtained results show that the personalization algorithm is able to capture the rules associated with the simulated concepts indicating its potential to be used in real-world settings.

In future studies, the aim is to validate the JITAI personalization mechanism empirically throughout the randomized controlled trial to be carried out in the scope of POWER2DM with 280 diabetes patients in total[107].

REFERENCES

[1] B. W. Ward, J. S. Schiller, and R. A. Goodman, "Peer reviewed: Multiple chronic conditions among us adults: A 2012 update," *Preventing chronic disease*, vol. 11, 2014.

[2] W. H. Organization, "Noncommunicable diseases." http://www.who.int/en/news-room/fact-sheets/detail/noncommunicable-diseases, Last accessed: 20.10.2018.

[3] L. Guariguata, D. R. Whiting, I. Hambleton, J. Beagley, U. Linnenkamp, and J. E. Shaw, "Global estimates of diabetes prevalence for 2013 and projections for 2035," *Diabetes research and clinical practice*, vol. 103, no. 2, pp. 137–149, 2014.

[4] R. A. Vigersky, L. Fish, P. Hogan, A. Stewart, S. Kutler, P. W. Ladenson, M. McDermott, and K. H. Hupart, "The clinical endocrinology workforce: current status and future projections of supply and demand," *The Journal of Clinical Endocrinology & Metabolism*, vol. 99, no. 9, pp. 3112–3121, 2014.

[5] C. C. Quinn, A. L. Gruber-Baldini, M. Shardell, K. Weed, S. S. Clough, M. Peeples, M. Terrin, L. Bronich-Hall, E. Barr, and D. Lender, "Mobile diabetes intervention study: testing a personalized treatment/behavioral communication intervention for blood glucose control," *Contemporary Clinical Trials*, vol. 30, no. 4, pp. 334–346, 2009.

[6] V. Iyengar, A. Wolf, A. Brown, and K. Close, "Challenges in diabetes care: Can digital health help address them?," *Clinical Diabetes*, vol. 34, no. 3, pp. 133–141, 2016.

[7] C. K. Roberts and R. J. Barnard, "Effects of exercise and diet on chronic disease," *Journal of applied physiology*, vol. 98, no. 1, pp. 3–30, 2005.

[8] M. Moattari, A. Ghobadi, P. Beigi, and G. Pishdad, "Impact of self management on metabolic control indicators of diabetes patients," *Journal of Diabetes & Metabolic Disorders*, vol. 11, no. 1, p. 6, 2012.

[9] M. M. Funnell and R. M. Anderson, "Empowerment and self-management of diabetes," *Clinical diabetes*, vol. 22, no. 3, pp. 123–127, 2004.

[10] D. Almirall, I. Nahum-Shani, N. E. Sherwood, and S. A. Murphy, "Introduction to smart designs for the development of adaptive interventions: with application to weight loss research," *Translational behavioral medicine*, vol. 4, no. 3, pp. 260–274, 2014.

[11] E. Deci and R. M. Ryan, *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media, 1985.

[12] H. Holmen, A. K. Wahl, M. C. Småstuen, and L. Ribu, "Tailored communication within mobile apps for diabetes self-management: a systematic review," *Journal of medical Internet research*, vol. 19, no. 6, 2017.

[13] M. Hood, R. Wilson, J. Corsica, L. Bradley, D. Chirinos, and A. Vivo, "What do we know about mobile applications for diabetes self-management? a review of reviews," *Journal of behavioral medicine*, vol. 39, no. 6, pp. 981–994, 2016.

[14] G. Kok, N. H. Gottlieb, G.-J. Y. Peters, P. D. Mullen, G. S. Parcel, R. A. Ruiter, M. E. Fernández, C. Markham, and L. K. Bartholomew, "A taxonomy of behaviour change methods: an intervention mapping approach," *Health psychology review*, vol. 10, no. 3, pp. 297–312, 2016.

[15] L. K. B. Eldredge, C. M. Markham, R. A. Ruiter, G. Kok, M. E. Fernandez, and G. S. Parcel, *Planning health promotion programs: an intervention mapping approach*. John Wiley & Sons, 2016.

[16] I. Nahum-Shani, S. N. Smith, A. Tewari, K. Witkiewitz, L. M. Collins, B. Spring, and S. Murphy, "Just in time adaptive interventions (jitais): An organizing framework for ongoing health behavior support," *Methodology Center technical report*, vol. 2014, pp. 14–126, 2014.

[17] D. Spruijt-Metz and W. Nilsen, "Dynamic models of behavior for just-in-time adaptive interventions," *IEEE Pervasive Computing*, vol. 13, no. 3, pp. 13–17, 2014.

[18] I. Nahum-Shani, S. N. Smith, B. J. Spring, L. M. Collins, K. Witkiewitz, A. Tewari, and S. A. Murphy, "Just-in-time adaptive interventions (jitais) in mobile health: key components and design principles for ongoing health behavior support," *Annals of Behavioral Medicine*, vol. 52, no. 6, pp. 446–462, 2017.

[19] V. Pejovic and M. Musolesi, "Interruptme: designing intelligent prompting mechanisms for pervasive applications," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 897–908, ACM, 2014.

[20] C. Pop-Eleches, H. Thirumurthy, J. P. Habyarimana, J. G. Zivin, M. P. Goldstein, D. De Walque, L. Mackeen, J. Haberer, S. Kimaiyo, J. Sidle, *et al.*, "Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders," *AIDS (London, England)*, vol. 25, no. 6, p. 825, 2011.

[21] I.-J. Chen and C.-C. Chang, "Content presentation modes in mobile language listening tasks: English proficiency as a moderator," *Computer Assisted Language Learning*, vol. 24, no. 5, pp. 451–470, 2011.

[22] C. A. Pellegrini, A. F. Pfammatter, D. E. Conroy, and B. Spring, "Smartphone applications to support weight loss: current perspectives," *Advanced health care technologies*, vol. 1, p. 13, 2015.

[23] G. Elwyn, D. Frosch, R. Thomson, N. Joseph-Williams, A. Lloyd, P. Kinnersley, E. Cording, D. Tomson, C. Dodd, S. Rollnick, *et al.*, "Shared decision making: a model for clinical practice," *Journal of general internal medicine*, vol. 27, no. 10, pp. 1361–1367, 2012.

[24] R. S. Sutton, A. G. Barto, *et al.*, *Reinforcement learning: An introduction*. MIT press, 1998.

[25] V. Pejovic and M. Musolesi, “Anticipatory mobile computing for behaviour change interventions,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pp. 1025–1034, ACM, 2014.

[26] A. Lazaric, “Transfer in reinforcement learning: a framework and a survey,” in *Reinforcement Learning*, pp. 143–173, Springer, 2012.

[27] M. E. Taylor, P. Stone, and Y. Liu, “Transfer learning via inter-task mappings for temporal difference learning,” *Journal of Machine Learning Research*, vol. 8, no. Sep, pp. 2125–2167, 2007.

[28] P. Kormushev, K. Nomoto, F. Dong, and K. Hirota, “Time hopping technique for faster reinforcement learning in simulations,” *CYBERNETICS AND INFORMATION TECHNOLOGIES*, vol. 11, no. 3, 2011.

[29] C. J. Watkins and P. Dayan, “Q-learning,” *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[30] R. S. Sutton, A. G. Barto, *et al.*, *Reinforcement learning: An introduction*. MIT press, 1998.

[31] L. Torrey and J. Shavlik, “Transfer learning,” in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, pp. 242–264, IGI Global, 2010.

[32] D. Ben-Zeev, C. J. Brenner, M. Begale, J. Duffecy, D. C. Mohr, and K. T. Mueser, “Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia,” *Schizophrenia bulletin*, vol. 40, no. 6, pp. 1244–1253, 2014.

[33] A. Fioravanti, G. Fico, D. Salvi, R. I. García-Betances, and M. T. Arredondo, “Automatic messaging for improving patients engagement in diabetes management: an exploratory study,” *Medical & biological engineering & computing*, vol. 53, no. 12, pp. 1285–1294, 2015.

[34] B. Y. Laing, C. M. Mangione, C.-H. Tseng, M. Leng, E. Vaisberg, M. Mahida, M. Bholat, E. Glazier, D. E. Morisky, and D. S. Bell, “Effectiveness of a

smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial," *Annals of internal medicine*, vol. 161, no. 10_Supplement, pp. S5–S12, 2014.

[35] D. H. Gustafson, F. M. McTavish, M.-Y. Chih, A. K. Atwood, R. A. Johnson, M. G. Boyle, M. S. Levy, H. Driscoll, S. M. Chisholm, L. Dillenburg, *et al.*, "A smartphone application to support recovery from alcoholism: a randomized clinical trial," *JAMA psychiatry*, vol. 71, no. 5, pp. 566–572, 2014.

[36] S. Dantzig, G. Geleijnse, and A. T. Halteren, "Toward a persuasive mobile application to reduce sedentary behavior," *Personal and ubiquitous computing*, vol. 17, no. 6, pp. 1237–1246, 2013.

[37] H. E. Payne, C. Lister, J. H. West, and J. M. Bernhardt, "Behavioral functionality of mobile apps in health interventions: a systematic review of the literature," *JMIR mHealth and uHealth*, vol. 3, no. 1, 2015.

[38] J. Nicholas, M. E. Larsen, J. Proudfoot, and H. Christensen, "Mobile apps for bipolar disorder: a systematic review of features and content quality," *Journal of medical Internet research*, vol. 17, no. 8, 2015.

[39] J. S. Baron, S. Hirani, and S. P. Newman, "A randomised, controlled trial of the effects of a mobile telehealth intervention on clinical and patient-reported outcomes in people with poorly controlled diabetes," *Journal of telemedicine and telecare*, vol. 23, no. 2, pp. 207–216, 2017.

[40] W. C. Hsu, K. H. K. Lau, R. Huang, S. Ghiloni, H. Le, S. Gilroy, M. Abrahamson, and J. Moore, "Utilization of a cloud-based diabetes management program for insulin initiation and titration enables collaborative decision making between healthcare providers and patients," *Diabetes technology & therapeutics*, vol. 18, no. 2, pp. 59–67, 2016.

[41] A. D. Association *et al.*, "5. prevention or delay of type 2 diabetes," *Diabetes Care*, vol. 38, no. Supplement 1, pp. S31–S32, 2015.

[42] K. Waki, H. Fujita, Y. Uchimura, K. Omae, E. Aramaki, S. Kato, H. Lee, H. Kobayashi, T. Kadowaki, and K. Ohe, "Dialbetics: a novel smartphone-

based self-management support system for type 2 diabetes patients," *Journal of diabetes science and technology*, vol. 8, no. 2, pp. 209–215, 2014.

[43] J. Diabetes Society, "Treatment guide for diabetes." http://www.fa.kyorin.co.jp/jds/uploads/Treatment_Guide_for_Diabetes_2016-2017.pdf, Last accessed: 19.10.2018.

[44] E. D. Bateman, S. Hurd, P. Barnes, J. Bousquet, J. Drazen, M. FitzGerald, P. Gibson, K. Ohta, P. O'byrne, S. Pedersen, *et al.*, "Global strategy for asthma management and prevention: Gina executive summary," *European Respiratory Journal*, vol. 31, no. 1, pp. 143–178, 2008.

[45] W.-T. Liu, C.-D. Huang, C.-H. Wang, K.-Y. Lee, S.-M. Lin, and H.-P. Kuo, "A mobile telephone-based interactive self-care system improves asthma control," *European Respiratory Journal*, vol. 37, no. 2, pp. 310–317, 2011.

[46] D. Spruijt-Metz, E. Hekler, N. Saranummi, S. Intille, I. Korhonen, W. Nilsen, D. E. Rivera, B. Spring, S. Michie, D. A. Asch, *et al.*, "Building new computational models to support health behavior change and maintenance: new opportunities in behavioral research," *Translational behavioral medicine*, vol. 5, no. 3, pp. 335–346, 2015.

[47] J.-E. Navarro-Barrientos, D. E. Rivera, and L. M. Collins, "A dynamical model for describing behavioural interventions for weight loss and body composition change," *Mathematical and computer modelling of dynamical systems*, vol. 17, no. 2, pp. 183–203, 2011.

[48] R. A. Hammond, J. T. Ornstein, L. K. Fellows, L. Dubé, R. Levitan, and A. Dagher, "A model of food reward learning with dynamic reward exposure," *Frontiers in computational neuroscience*, vol. 6, p. 82, 2012.

[49] M.-Y. Chih, T. Patton, F. M. McTavish, A. J. Isham, C. L. Judkins-Fisher, A. K. Atwood, and D. H. Gustafson, "Predictive modeling of addiction lapses in a mobile health application," *Journal of substance abuse treatment*, vol. 46, no. 1, pp. 29–35, 2014.

[50] S. Mohan, A. Venkatakrishnan, M. Silva, and P. Pirolli, "On designing a social coach to promote regular aerobic exercise.," in *AAAI*, pp. 4721–4727, 2017.

[51] S. P. Goldstein, B. C. Evans, D. Flack, A. Juarascio, S. Manasse, F. Zhang, and E. M. Forman, "Return of the jitai: applying a just-in-time adaptive intervention framework to the development of m-health solutions for addictive behaviors," *International journal of behavioral medicine*, vol. 24, no. 5, pp. 673–682, 2017.

[52] W. T. Riley, A. Cesar, D. E. Rivera, *et al.*, "The importance of behavior theory in control system modeling of physical activity sensor data," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pp. 6880–6883, IEEE, 2014.

[53] I. Ajzen, "From intentions to actions: A theory of planned behavior," in *Action control*, pp. 11–39, Springer, 1985.

[54] A. Bandura, "Social foundations of thought and action," *Englewood Cliffs, NJ*, vol. 1986, 1986.

[55] C. A. Martin, D. E. Rivera, W. T. Riley, E. B. Hekler, M. P. Buman, M. A. Adams, and A. C. King, "A dynamical systems model of social cognitive theory," in *American Control Conference (ACC), 2014*, pp. 2407–2412, IEEE, 2014.

[56] E. W. Boyer, R. Fletcher, R. J. Fay, D. Smelson, D. Ziedonis, and R. W. Picard, "Preliminary efforts directed toward the detection of craving of illicit substances: the iheal project," *Journal of Medical Toxicology*, vol. 8, no. 1, pp. 5–9, 2012.

[57] L. G. Morrison, C. Hargood, V. Pejovic, A. W. Geraghty, S. Lloyd, N. Goodman, D. T. Michaelides, A. Weston, M. Musolesi, M. J. Weal, *et al.*, "The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: An exploratory trial," *PloS one*, vol. 12, no. 1, p. e0169162, 2017.

[58] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, (New York, NY, USA), pp. 785–794, ACM, 2016.

[59] M. Pielot, B. Cardoso, K. Katevas, J. Serrà, A. Matic, and N. Oliver, "Beyond interruptibility: Predicting opportune moments to engage mobile phone users," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, pp. 91:1–91:25, Sept. 2017.

[60] H. Oh, L. Jalali, R. Jain, *et al.*, "An intelligent notification system using context from real-time personal activity monitoring," in *2015 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, IEEE, 2015.

[61] A. Mehrotra, M. Musolesi, R. Hendley, and V. Pejovic, "Designing content-driven intelligent notification mechanisms for mobile applications," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 813–824, ACM, 2015.

[62] A. Mehrotra, V. Pejovic, J. Vermeulen, R. Hendley, and M. Musolesi, "My phone and me: understanding people's receptivity to mobile notifications," in *Proceedings of the 2016 CHI conference on human factors in computing systems*, pp. 1021–1032, ACM, 2016.

[63] J. E. Fischer, C. Greenhalgh, and S. Benford, "Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications," in *Proceedings of the 13th international conference on human computer interaction with mobile devices and services*, pp. 181–190, ACM, 2011.

[64] J. Kelly, P. Gooding, D. Pratt, J. Ainsworth, M. Welford, and N. Tarrier, "Intelligent real-time therapy: Harnessing the power of machine learning to optimise the delivery of momentary cognitive–behavioural interventions," *Journal of Mental Health*, vol. 21, no. 4, pp. 404–414, 2012.

[65] H. Lei, A. Tewari, and S. Murphy, "An actor-critic contextual bandit algorithm for personalized interventions using mobile devices," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[66] H. Lei, A. Tewari, and S. A. Murphy, "An actor-critic contextual bandit algorithm for personalized mobile health interventions," *arXiv preprint arXiv:1706.09090*, 2017.

[67] J. Langford and T. Zhang, "The epoch-greedy algorithm for multi-armed bandits with side information," in *Advances in neural information processing systems*, pp. 817–824, 2008.

[68] C. J. C. H. Watkins, *Learning from delayed rewards*. PhD thesis, King's College, Cambridge, 1989.

[69] M. K. Bloch, "Temporal second difference traces," *arXiv preprint arXiv:1104.4664*, 2011.

[70] H. van Seijen and S. Whiteson, "Postponed updates for temporal-difference reinforcement learning," in *Intelligent Systems Design and Applications, 2009. ISDA'09. Ninth International Conference on*, pp. 665–672, IEEE, 2009.

[71] M. E. Taylor and P. Stone, "Transfer learning for reinforcement learning domains: A survey," *Journal of Machine Learning Research*, vol. 10, no. Jul, pp. 1633–1685, 2009.

[72] A. Lazaric, *Knowledge transfer in reinforcement learning*. PhD thesis, PhD thesis, Politecnico di Milano, 2008.

[73] E. Pakizeh, M. M. Pedram, and M. Palhang, "Multi-criteria expertness based cooperative method for sarsa and eligibility trace algorithms," *Applied Intelligence*, vol. 43, no. 3, pp. 487–498, 2015.

[74] P. Van de Ven, H. O'Brien, R. Henriques, M. Klein, R. Msetfi, J. Nelson, A. Rocha, J. Ruwaard, D. O'Sullivan, H. Riper, *et al.*, "Ultemat: A mobile framework for smart ecological momentary assessments and interventions," *Internet Interventions*, vol. 9, pp. 74–81, 2017.

[75] S. Shiffman, A. A. Stone, and M. R. Hufford, "Ecological momentary assessment," *Annu. Rev. Clin. Psychol.*, vol. 4, pp. 1–32, 2008.

[76] S. Pandey, W. Voorsluys, S. Niu, A. Khandoker, and R. Buyya, "An autonomic cloud environment for hosting ecg data analysis services," *Future Generation Computer Systems*, vol. 28, no. 1, pp. 147–154, 2012.

[77] C. Doukas and I. Maglogiannis, "Managing wearable sensor data through

cloud computing," in *Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on*, pp. 440–445, IEEE, 2011.

[78] G. Fortino, M. Pathan, and G. Di Fatta, "Bodycloud: Integration of cloud computing and body sensor networks," in *Cloud Computing Technology and Science (CloudCom), 2012 IEEE 4th International Conference on*, pp. 851–856, IEEE, 2012.

[79] M. Al Hemairy, M. A. Serhani, S. Amin, and M. Al Ahmed, "Integrated and scalable architecture for providing cost-effective remote health monitoring," in *Developments in eSystems Engineering (DeSE), 2016 9th International Conference on*, pp. 74–80, IEEE, 2016.

[80] C. Doukas and I. Maglogiannis, "Bringing iot and cloud computing towards pervasive healthcare," in *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012 Sixth International Conference on*, pp. 922–926, IEEE, 2012.

[81] J. Burt, J. Rick, T. Blakeman, J. Protheroe, M. Roland, and P. Bower, "Care plans and care planning in long-term conditions: a conceptual model," *Primary health care research & development*, vol. 15, no. 4, pp. 342–354, 2014.

[82] V. Dolk, "Survey reinforcement learning," *Eindhoven University of Technology*, 2010.

[83] P. Lally, C. H. Van Jaarsveld, H. W. Potts, and J. Wardle, "How are habits formed: Modelling habit formation in the real world," *European journal of social psychology*, vol. 40, no. 6, pp. 998–1009, 2010.

[84] R. Tobias, "Changing behavior by memory aids: a social psychological model of prospective memory and habit development tested with dynamic field data.," *Psychological review*, vol. 116, no. 2, p. 408, 2009.

[85] S. Gonul, T. Namli, S. Huisman, G. B. Laleci Erturkmen, I. H. Toroslu, and A. Cosar, "An expandable approach for design and personalization of digital, just-in-time adaptive interventions," *Journal of the American Medical Informatics Association*, 2018. (Accepted for publication).

[86] J. A. Gregg, G. M. Callaghan, and S. C. Hayes, *The diabetes lifestyle book: Facing your fears and making changes for a long and healthy life*. New Harbinger Publications, 2007.

[87] S. Michie, M. Richardson, M. Johnston, C. Abraham, J. Francis, W. Hardeman, M. P. Eccles, J. Cane, and C. E. Wood, "The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions," *Annals of behavioral medicine*, vol. 46, no. 1, pp. 81–95, 2013.

[88] S. Michie, S. Ashford, F. F. Sniehotta, S. U. Dombrowski, A. Bishop, and D. P. French, "A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the calo-re taxonomy," *Psychology & health*, vol. 26, no. 11, pp. 1479–1498, 2011.

[89] M. Plößnig, I. Smith, S. Huisman, J. Sont, T. Namlı, S. Gönül, H. Kroon, A. De Graaf, L. Vogt, F. Strohmeier, P. van Empelen, H. van Keulen, and W. Otten, "Power2dm - d3.1.1 dynamic behaviour change intervention models for self-management." http://www.power2dm.eu/wp-content/uploads/Power2DM-D3.1-1.pdf, Last accessed: 29.10.2018.

[90] J. W. Backus, "The syntax and semantics of the proposed international algebraic language of the zurich acm-gamm conference," *Proceedings of the International Comference on Information Processing, 1959*, 1959.

[91] S. Gonul, T. Namli, M. Baskaya, A. A. Sinaci, A. Cosar, and I. H. Toroslu, "Optimization of just-in-time adaptive interventions using reinforcement learning," in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp. 334–341, Springer, 2018.

[92] M. W. Floyd and B. Esfandiari, "Supplemental observation acquisition for learning by observation agents," *Applied Intelligence*, pp. 1–17, 2018.

[93] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[94] X. Meng, J. Bradley, B. Yavuz, E. Sparks, S. Venkataraman, D. Liu, J. Freeman, D. Tsai, M. Amde, S. Owen, *et al.*, "Mllib: Machine learning in apache

spark," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 1235–1241, 2016.

[95] S. A. Mulvaney, L. M. Ritterband, and L. Bosslet, "Mobile intervention design in diabetes: review and recommendations," *Current diabetes reports*, vol. 11, no. 6, p. 486, 2011.

[96] E. Bainomugisha, A. L. Carreton, T. v. Cutsem, S. Mostinckx, and W. d. Meuter, "A survey on reactive programming," *ACM Computing Surveys (CSUR)*, vol. 45, no. 4, p. 52, 2013.

[97] N. Marz and J. Warren, *Big Data: Principles and best practices of scalable real-time data systems*. New York; Manning Publications Co., 2015.

[98] A. D. Association *et al.*, "5. prevention or delay of type 2 diabetes," *Diabetes Care*, vol. 40, no. Supplement 1, pp. S44–S47, 2017.

[99] J. DIABETES CENTER & JOSLIN CLINIC *et al.*, "Clinical guideline for adults with diabetes." https://www.joslin.org/docs/CLINICAL-GUIDELINE-FOR-ADULTS-WITH-DIABETES-Rev-05-17-2017.pdf, Last accessed: 16.10.2018.

[100] K. Rockwood, X. Song, C. MacKnight, H. Bergman, D. B. Hogan, I. McDowell, and A. Mitnitski, "A global clinical measure of fitness and frailty in elderly people," *Canadian Medical Association Journal*, vol. 173, no. 5, pp. 489–495, 2005.

[101] L. Meneghini, C. Koenen, W. Weng, and J.-L. Selam, "The usage of a simplified self-titration dosing guideline (303 algorithm) for insulin detemir in patients with type 2 diabetes–results of the randomized, controlled predictive$^{\mathrm{TM}}$ 303 study," *Diabetes, Obesity and Metabolism*, vol. 9, no. 6, pp. 902–913, 2007.

[102] H. L. Hom and F. R. Maxwell, "The impact of task difficulty expectations on intrinsic motivation," *Motivation and Emotion*, vol. 7, no. 1, pp. 19–24, 1983.

[103] D. Scasserra, "The influence of perceived task difficulty on task performance," 2008.

[104] F. Morrison, M. Shubina, and A. Turchin, “Encounter frequency and serum glucose level, blood pressure, and cholesterol level control in patients with diabetes mellitus,” *Archives of internal medicine*, vol. 171, no. 17, pp. 1542–1550, 2011.

[105] G. A. Rummery and M. Niranjan, *On-line Q-learning using connectionist systems*, vol. 37. University of Cambridge, Department of Engineering Cambridge, England, 1994.

[106] P. Klasnja, E. B. Hekler, S. Shiffman, A. Boruvka, D. Almirall, A. Tewari, and S. A. Murphy, “Microrandomized trials: An experimental design for developing just-in-time adaptive interventions.,” *Health Psychology*, vol. 34, no. S, p. 1220, 2015.

[107] J. Sont, I. Smith, B. Uitbeijerse, W. Heemstra-Klein Wassink, S. Huisman, E. de Koning, J. Snoeck-Stroband, J. Delgado, and A. de Graaf, “Power2dm - d5.2.2 evaluation of campaign methodology.” http://www.power2dm.eu/wp-content/uploads/Power2DM-D5.3.pdf, Last accessed: 18.10.2018.

[108] C. M. Van der Feltz-Cornelis, J. Nuyen, C. Stoop, J. Chan, A. M. Jacobson, W. Katon, F. Snoek, and N. Sartorius, “Effect of interventions for major depressive disorder and significant depressive symptoms in patients with diabetes mellitus: a systematic review and meta-analysis,” *General hospital psychiatry*, vol. 32, no. 4, pp. 380–395, 2010.

APPENDIX A

BUILT-IN CONSTRUCTS OF RULE DEFINITION LANGUAGE

A.1 Tailoring Variables

Tailoring variables are data integration and processing modules that respectively provide the external information to the rest of the system and perform custom analytics on the integrated data. Below, the built-in contexts included in the current version of the *Rule Definition Language* along with their brief descriptions are presented.

- **goal:** How far the person has reached the default goal associated to a behavior. This gets one of the enumerated values defined below:
  - Not achieved the goal (0)
  - Almost achieved goal (1)
  - About to achieve goal (2)
  - Achieved goal (3)
  - Achieved more than goal (4)
- **adherence:** Adherence to the behavior. Takes values between 0 and 1, which means no adherence and complete adherence respectively. A behavior may have different goals, so may have different adherence calculations. Some of the examples used in the proposed system are shown below:
  - **adherence: nmeds:** Performance calculated according to the number of medication intakes matched.
  - **adherence: duration:** Performance calculated according to the duration of the physical exercise and what is planned.

  - **adherence: logging:** Performance calculated if the planned meals are logged succesfully.
  - **adherence: lowcarb:** Performance calculated based on the carb intake and what is planned.

- **steps:** Step counts.
- **stress_level:** Enumerated stress values obtained through a 5-valued likert scale where the values range from *none* to *very high*
- **time_to_threshold:** Indicates the number of steps to reach a pre-defined performance level.

A.2 Temporals

Temporal constructs are applied to the contexts to obtain values for a specific period, peak points or combination of both.

- **daily:** Daily average value for a given context.
- **weekly:** Weekly average value for a given context.
- **monthly:** Monthly average value for a given context.
- **best:** Best value for a given context.
- **worst:** Worst value for a given context.

Period-based and peak values can be combined as *best-day, best-week, best-month, worst-day, worst-week* and *worst-month*.

A.3 Placeholders

Placeholders are custom data processing modules of which values can be embedded in intervention content. They enable designing generic intervention option templates specific to a Behavioral Change Technique (BCT) such that placeholders could be instantiated with a dynamically calculated value based on the person-specific data. The followings are the currently available placeholders:

- **streak_value:** Calculates the number of successive achievement of a goal. It is used together with the *streak_temporal* placeholder and the calculation is done based on the temporal constraint specified in *streak_temporal* placeholder.
- **streak_temporal:** Temporal period for which the goal is reached. It can take *months, weeks, days* and *times* values.
- **goal_temporal:** Temporal period associated with a goal.
- **goal_remaining:** Number of remaining goals.
- **action_time:** Planned time for the activity specified in the action plan.

The rest of the placeholders are related to comparison of the last goal performance of the person with either the average of the previous performances of the person him/herself or performance of the rest of the population.

- **comparison_temporal:** Temporal period for the comparison.
- **comparison_value:** The difference between the last value of a goal and compared value.
- **comparison_population_percentage:** Percentage of the population, which the person is better than considering the goal performance.
- **comparison_population_number:** Number of people, which the person is better than concerning the goal performance.

The following 4 placeholders are also related to comparison of the current goal performance. But this time, the calculations are performed assuming that the person performs the behavior in the next opportunity. The aim is to motivate people by showing how the goal performance improves when the behavior is performed.

- **comparison_simulation_temporal:**

- **comparison_simulation_value:**
- **comparison_simulation_population_percentage:**
- **comparison_simulation_population_number:**

A.4 Temporal Index

Indices allow data retrieval for contexts for a specific period in the past.

- If no index is associated with a context, the latest value of the context is considered
  - e.g. goal = ACHIEVED means, if the person reaches the latest goal
  - e.g. goal.daily = ACHIEVED means, if the person reaches the goal today
  - e.g. goal.monthly = ACHIEVED means, if the person reaches the goal this month
- If index is positive, it means the comparison should be satisfied for each of the latest values specified by index
  - e.g. goal[4] = NOT ACHIEVED means, if the person does not achieve his/her goals for latest 4 actions
  - e.g. stress[7] > NORMAL means, for the last 7 stress recordings, the person has higher-than-normal stress levels (i.e. HIGH or VERY HIGH)
  - e.g. goal.daily[3] = ACHIEVED means, if the person achieves his/her daily goals for the last 3 days
- If index is negative, it means the comparison should be satisfied for the temporal given by the index itself
  - goal.weekly[-1] = NOT ACHIEVEDmeans, if the person does not achieve his/her last week goal
  - stress.monthly<stress.monthly[-2] means, if the average stress of the person's stress in this month is less than 2 months-before average

A.5 Logical Operators

Rule conditions can be joined by “and”s; where both of them should be satisfied

- e.g. goal.monthly = 2 and goal.montly[-1] = 2 means if the person reaches his/her goal both this month and last month

Rules can also be specified sequentially in a comma-separated way. In this case, they are evaluated sequentially. Considering the following example, first, the monthly goal will be checked and then the weekly and finally the daily. The subsequent rules are tested only if the former ones fail.

- e.g. ["goal.monthly = 3", "goal.weekly = 3","goal.daily = 3"]

A.6 Behavior Change Techniques

Communication Engine currently implements 4 literature-driven behavior change techniques as listed below:

- General reinforcement
- Positive comparison with self
- Positive comparison with others
- Goal / action plan monitoring

APPENDIX B

ANALYSIS OF CALO-RE TAXONOMY

Table B.1: Implementation feasibility for the BCTs of CALO-RE taxonomy

BCT number in the taxonomy	**BCT Name**	**Implementation Feasibility**
1	Provide information on consequences of behaviour in general	Yes
2	Provide information on consequences of behaviour to the individual	Yes
4	Provide normative information about others' behaviour	Yes
5	Goal setting (behaviour)	Yes
6	Goal setting (outcome)	Yes
7	Action planning	Yes
8	Barrier identification/Problem solving	Yes
9	Set graded tasks	Yes
10	Prompt review of behavioural goals	Yes
11	Prompt review of outcome goals	Yes
12	Prompt rewards contingent on effort or progress towards behaviour	Yes
13	Provide rewards contingent on successful behaviour	Yes

Continued on next page

Table B.1 – *Continued from previous page*

14	Shaping	Yes
16	Prompt self-monitoring of behaviour	Yes
17	Prompt self-monitoring of behavioural outcome	Yes
18	Prompting focus on past success	Yes
19	Provide feedback on performance	Yes
20	Provide information on where and when to perform the behaviour	Yes
21	Provide instruction on how to perform the behaviour	Yes
22	Model/ Demonstrate the behaviour	Yes
23	Teach to use prompts/ cues	Yes
26	Prompt practice	Yes
27	Use of follow up prompts	Yes
28	Facilitate social comparison	Yes
31	Prompt anticipated regret	Yes
32	Fear Arousal	Yes
33	Prompt Self talk	Yes
34	Prompt use of imagery	Yes
35	Relapse prevention/ Coping planning	Yes
36	Stress management/Emotional control training	Yes
38	Time management	Yes
3	Provide information about others' approval	No
15	Prompting generalization of a target behaviour	No
24	Environmental restructuring	No
25	Agree behavioural contract	No
29	Plan social support/ social change	No

Continued on next page

Table B.1 – *Continued from previous page*

30	Prompt identification as role model/ position advocate	No
37	Motivational interviewing	No
39	General communication skills training	No
40	Stimulate anticipation of future rewards	No

APPENDIX C

EXAMPLE INTERVENTION DEFINITIONS DRIVEN BY THE CALO-RE TAXONOMY

Two intervention definition examples are presented corresponding to *set graded tasks* and *relapse prevention / coping planning* techniques introduced in the CALO-RE taxonomy. For each example, a set of properties are provided including: a *description* explaining the purpose of the intervention; *decision points* specifying the moments when the interventions would be considered for delivery; *behavior* targeted by the intervention; *behavior change technique (BCT)* deriving the *content* of the intervention and *decision rules* that should be satisfied for activating the intervention delivery at the specified decision points.

Table C.1: Intervention example implementing the *setting grated tasks* technique

Template Element	**Value**
description	The person complied with the physical exercise related activities concerning the scheduled timings. However, s/he has not fulfilled the activities in terms of intensity. So, the system suggests milder activities.
decision points	event = [post_action]
behavior	Physical exercise
BCT	Setting graded tasks

Continued on next page

Table C.1 – *Continued from previous page*

decision rules	adherence.weekly> 0.8 and adherence:intensity.weekly< 0.5
content	• en: You are on track with your physical exercises with an ${*weekly_action_plan_adherence*}. However, it seems it would be better if we lower the intensity of the activities. How about ${*milder_physica_exercise_amount*} minutes each day?
associated goal	• 30 minutes of physical activity at least 4 days in a week • 3 kg reduction in body weight in the next 6 months

In case patients have both diabetes and depression, both diseases require simultaneous attention[108]. In the following example, it is inconclusively assumed that increasing levels of stress and irregular carbohydrate intakes in the last 3 days might be an indicator of depression.

Table C.2: Intervention example implementing the *relapse prevention/coping planning* technique

Template Element	**Value**
description	The system first checks that the person has complied with both the carbohydrate intake and stress level loggings. Then, it detects that the person has increasing levels of stress and irregular levels of carbohydrate intake considering the data aggregated during the last 3 days.

Continued on next page

Table C.2 – *Continued from previous page*

decision points	event = [post_action]
behavior	Carbohydrate monitoring
BCT	Relapse prevention/ coping planning
decision rules	adherence:logging.daily [3] > 0.8 and stress_level.daily[-2] >stress_level.daily [-3] and stress_level.daily [-1] >stress_level.daily [-2] and stress_level.daily [-1] >stress_level.daily and carbohydrate_intake(kcal)_variance.daily [3] > 300
content	• en: An change in the daily routine will help you nowadays. Would you consider scheduling an event with your friends?
associated goal	• Prevent lapses of depression

APPENDIX D

EXAMPLE INTERVENTION DEFINITIONS DRIVEN BY THE RESOURCES AVAILABLE IN LITERATURE

The first example, described in D.1 is related to recommendations given by the American Diabetes Association on prolonged sitting cases.

Table D.1: Intervention example encouring a person to take a break after 30 minutes of prolonged sitting

Template Element	**Value**
description	The person is sitting during the last 30 minutes and the system encourages him/her to take a break and walk.
decision points	event = [change in tailoring variable representing being sedentary]
behavior	Walking
BCT	Provide information on consequences of behavior to the individual
decision rules	retrospective_inactivity.minute[30]= 1
content	• en: Taking a break at every 30 minutes help you to cope with the diabetes and it's already 30 minutes you have been sitting. How about to take a break now?

Continued on next page

Table D.1 – *Continued from previous page*

associated goal	• Reducing HbA1c levels below 7% in the next 6 month • Prevent prolonged sedentary behavior during the day

The second example is driven by the Predictive 303 algorithm for adjusting the insulin levels based on the fasting plasma glucose measurements of the patient.

Table D.2: Intervention example adjusting the insulin dosage based on the Predictive 303 algorithm

Template Element	**Value**
description	Predictive 303 algorithm introduces a rule set for adjusting insulin detemir every 3 days based on the mean of three adjusted fasting plasma glucose (aFPG) as follows: if mean aFPG< 80 mg/dl, reduce dose by 3 unit; if aFPGis between 80 and 110 mg/dl, no change; and if aFPG> 110 mg/dl, increase dose by 3 units. Based on these rules, the system guides the person on adjusting his/her insulin intake levels. This specific intervention handles the first case.
decision points	event = [change in tailoring variable representing the aFBG measurements in the last 3 days]
behavior	Insulin intake
BCT	-
decision rules	aFPG.daily [3]<80

Continued on next page

Table D.2 – *Continued from previous page*

content	• en: You have 70 mg/dl of aFBG in average in the last 3 days. You should decrease the increase by 3 units to ${new_insulin_level}.
associated goal	• Prevent hyperglicemia lapses

APPENDIX E

EXAMPLE INTERVENTION DEFINITIONS FROM THE POWER2DM REAL-WORLD CASE STUDY

One example intervention definition is presented for each of the four behaviors addressed in the POWER2DM Project, namely blood glucose monitoring, carbohydrate monitoring, physical exercise and medication adherence. There are 120 intervention definitions in total addressing these behaviors. The definitions that are not presented in this study have been defined as reminders or motivations with a specific BCT considering the goal achievement. For example, the intervention described in E.1 is a reminder implementing positive comparison with others technique for blood glucose monitoring. It is supposed be delivered when the person is close achieving the associated goal. There could be multiple implementations of the same type of intervention that are valid in the same set of conditions.

Table E.1: Intervention example for blood glucose monitoring behavior

Template Element	**Value**
description	Patient has upcoming BG monitoring action and is close to achieve his monthly, weekly or daily goal, we remind him with motivation with a simulation comparison with others.
decision points	event = [upcoming_action]
behavior	Blood glucose monitoring
BCT	Positive comparison with others

Continued on next page

Table E.1 – *Continued from previous page*

decision rules	goal.monthly = 2", "goal.weekly = 2
content	• en: Just to remind you; You have an upcoming ${action_name} schedule(${action_time})! If you can complete it, your performance will be better than ${comparison_simulation_population_percentage}% of others ${comparison_simulation_temporal}. • es: Solo para recordarle; ¡Tiene una próxima ${action_name} programada (${action_time})! Si puede completarla, su rendimiento será mejor que ${comparison_simulation_population_percentage}% de los demás ${comparison_simulation_temporal}. • nl: Om je te helpen herinneren; Je heb teen schema voor ${action_name} (${action_time})! Als je het kun voltooien, is je prestatiemeter dan ${comparison_simulation_population_percentage}% van anderen ${comparison_simulation_temporal}.
associated goal	• Reducing HbA1c levels below 7% in the next 6 month • Monitoring blood glucose levels 3 times a day

Table E.2: Intervention example for carbohydrate monitoring behavior

Template Element	**Value**
description	Patient gets high carb in his last meal, but he was not that bad in the last two days, the system motivates her/him by comparing possible performance (if s/he can complete remaining tasks) with past.
decision points	event = [post_action]
behavior	Carbohydrate monitoring
BCT	Positive comparison with self
decision rules	adherence:logging = 1 and adherence:lowcarb< 1 and goal.daily = NOT ACHIEVED and goal.daily[2] != NOT ACHIEVED

Continued on next page

Table E.2 – *Continued from previous page*

content	• en: It seems, you didn't achieve low carbohydrate intake for your last meal. No problem, if you can adhere your remaining carb monitoring tasks, still you will be ${comparisonhsimulation_value}% better than ${comparison_simulation_temporal}. • es: Parece que no logró una ingesta bajaen carbohidratos en su última comida. No hay problema, si puede cumplir sus tareas de registro de carbohidratos restantes, aún así tendrá ${comparison_simulation_value}%mejor que ${comparison_simulation_temporal}. • nl: Het lijkteropdat je jedoel om weinig koolhydratente eten bij de laatste maaltijd niet hebt gehaald. Geen probleem, als je de resterende doelen/taken voor vandaag behaalt, dan ben je nog altijd ${comparison_simulation_value}% beter dan ${comparison_simulation_temporal}.
associated goal	• Reducing HbA1c levels below 7% in the next 6 month • Keep daily calorie intake under 2000 (kCal)

Table E.3: Intervention example for physical exercise behavior

Template Element	**Value**
description	Patient has almost achieved his last exercise goal, so we motivate him by general reinforcement.

Continued on next page

Table E.3 – *Continued from previous page*

decision points	event = [post_action]
behavior	Physical exercise
BCT	Generic reminder
decision rules	goal = ALMOST ACHIEVED and goal.weekly = NOT ACHIEVED
content	• en: Almost there for the ${goal_temporal} exercise session! Perform your exercise ${goal_remaining} minutes more than next time, and you will reach your goal. • es: ¡Casilisto para la sesión de ejercicio de ${goal_temporal}! Realices ejercicio ${goal_remaining} unos minutos más que la próximavez y alcanzará su objetivo. • nl: Je hebt ${goal_temporal} bijna behaald! Nog ${goal_remaining} minuten meer dan vorige keeren dan zul je jedo elbehalen.
associated goal	• Reducing HbA1c levels below 7% in the next 6 month • 30 minutes of physical activity at least 4 days in a week

Table E.4: Intervention example for medication adherence behavior

Template Element	**Value**
description	Patient forgets to log his last medication schedule, but he was not that bad in the last two days, we will motivate him by comparing possible performance (if he can complete remaining tasks) with past.
decision points	event = [post_action]
behavior	Medication adherence
BCT	Positive comparison with self
decision rules	adherence:nmeds = 0 and goal.daily[-1] = ACHIEVED

Continued on next page

Table E.4 – *Continued from previous page*

content	• en: It seems, you forget to log your last medication intake '${action_name}' (${action_time}). Please log your medication intake (or the reason not taking it). If you can complete your remaining tasks, still you will be ${comparison_simulation_value}% better than ${comparison_simulation_temporal}. • es: Parece que se olvidó registrar suúltimatoma de medicación '${action_name}' (${action_time}). Registre la toma de sumedicación (o la razón por la que no lo toma). Si puede completar sus tareas restantes, aúnestará ${comparison_simulation_value}% mejor que ${comparison_simulation_temporal}. • nl: Het lijktalsof je vergeten bent om je medicatie in tevoeren ${action_name} (${action_time}). Vul je medicatie alsnog in (of de reden waarom je deze niet hebt ingenomen). Als je jeresterendedoelen/taken voor vandaag behaalt, dan ben je gemiddeld nog steeds ${comparison_simulation_value}%beter in je doel dan ${comparison_simulation_temporal}.
associated goal	• Reducing HbA1c levels below 7% in the next 6 month • Keep monthly medication adherence ratio over 80%

APPENDIX F

INTERVENTION DEFINITIONS FOR THE SIMULATED CASE STUDY

In this section, the 3 imaginary interventions used in the simulated case study are presented.

Table F.1: Ordinary reminder intervention example

Template Element	**Value**
description	Patient has upcoming BG monitoring action and is close to achieve his monthly, weekly or daily goal, we remind him with motivation with a simulation comparison with others.
decision points	event = [upcoming_action]
behavior	Blood glucose monitoring
BCT	Prompt self-monitoring of behavior
decision rules	- (*Interpretation: At each decision point associated with the reminder interventions*)
content	• en: Just to remind you; You have an upcoming ${action_name} schedule (${action_time})!.

Continued on next page

Table F.1 – *Continued from previous page*

associated goal	• Reducing HbA1c levels below 7% in the next 6 month • Monitoring blood glucose levels 3 times a day

Table F.2: Reminder intervention example using social comparison BCT

Template Element	**Value**
description	Patient has an upcoming blood glucose monitoring activity. So, the system reminds him / her to perform the measurement. Besides reminding, the intervention also aims to motivate the person by comparing his/her performance with the rest of the population.
decision points	event = [upcoming_action]
behavior	Blood glucose monitoring
BCT	Prompt self-monitoring of behavior, Facilitate social comparison
decision rules	goal.daily = NOT_ACHIEVED (*Interpretation: Today's goal of blood glucose monitoring has not been achieved*)
content	• en: You have not measured your blood glucose today! Your ${action_name} performance is already better than the ${population_percentage}% of participants. Keep going! You have an upcoming ${action_name} schedule (${action_time})!

Continued on next page

Table F.2 – *Continued from previous page*

associated goal	• Reducing HbA1c levels below 7% in the next 6 month • Monitoring blood glucose levels 3 times a day

Table F.3: Motivation intervention example praising the performance

Template Element	**Value**
description	Patient has performed a scheduled blood glucose monitoring activity. The system motivates him/her to create an incentive for performing the behavior next time.
decision points	event = [post_action]
behavior	Blood glucose monitoring
BCT	Praising the performed behavior
decision rules	planned_activity=MISSED and goal = ACHIEVED *(Interpretation: The recent planned blood glucose monitoring activity considering the current time has been performed)*
content	• en: You are doing very good in ${action_name}. Keep it up!"
associated goal	• Reducing HbA1c levels below 7% in the next 6 month • Monitoring blood glucose levels 3 times a day

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Gönül, Suat
Nationality: Turkish (TC)
Date and Place of Birth: 24.02.1987, Lüleburgaz
Marital Status: Married
Phone: 0 312 2101763
Web Site: www.suatgonul.com

EDUCATION

Degree	Institution	Year of Graduation
M.S.	Computer Engineering Department, METU	2012
B.S.	Computer Engineering Department, METU	2010
High School	Lüleburgaz Anadolu Lisesi	2005

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
2012-Present	Yazılım Araştırma ve Geliştirme ve Danışmanlık Ltd. Şti.	Senior software engineer, Researcher
2010-2011	Yazılım Araştırma ve Geliştirme ve Danışmanlık Ltd. Şti.	Software engineer, Researcher
2008-2009	Yazılım Araştırma ve Geliştirme ve Danışmanlık Ltd. Şti.	Part-time software developer
2007	Computer Engineering Department, METU	Student assistant

PUBLICATIONS

Journal Publications

1. Gonul, S., Namli, T., Toroslu, I. H & Cosar, A. (2018). A Reinforcement Learning Based Algorithm for Personalization of Digital, Just-In-Time, Adaptive Interventions. *Applied Intelligence*. (Submitted)

2. Gonul, S., Namli, Huisman, S., Laleci Erturkmen, G. B., Toroslu, I. H & Cosar, A. (2018). An Expandable Approach for Design and Personalization of Digital, Just-In-Time Adaptive Interventions. *Journal of the American Medical Informatics Association.* (Accepted for publication)

3. Pacaci, A., Gonul, S., Sinaci, A. A., Yuksel, M., & Erturkmen, G. B. L. (2018). A Semantic Transformation Methodology for the Secondary Use of Observational Healthcare Data in Postmarketing Safety Studies. *Frontiers in pharmacology, 9.*

4. Yuksel, M., Gonul, S., Laleci Erturkmen, G. B., Sinaci, A. A., Invernizzi, P., Facchinetti, S., ... & De Roo, J. (2016). An interoperability platform enabling reuse of electronic health records for signal verification studies. *BioMed research international, 2016.*

5. Sinaci, A. A., Laleci Erturkmen, G. B., Gonul, S., Yuksel, M., Invernizzi, P., Thakrar, B., ... & Cicekli, N. K. (2015). Postmarketing safety study tool: a web based, dynamic, and interoperable system for postmarketing drug surveillance studies. *BioMed research international, 2015.*

International Conference Publications

1. Gonul, S., Namli, T., Baskaya, M., Sinaci, A. A., Cosar, A., & Toroslu, I. H. (2018, June). Optimization of Just-in-Time Adaptive Interventions Using Reinforcement Learning. *In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (pp. 334-341). Springer, Cham.*

2. Innerbichler, J., Gonul, S., Damjanovic-Behrendt, V., Mandler, B., & Strohmeier, F. (2017, June). Nimble collaborative platform: Microservice architectural approach to federated iot. *In Global Internet of Things Summit (GIoTS), 2017 (pp. 1-6). IEEE.*

3. Deng, Q., Gonul, S., Kabak, Y., Gessa, N., Glachs, D., Gigante, F., Damjanovic-Behrendt, V., Hribernik K. & Klaus-Dieter, T.An Ontology Framework for Multisided Platform Interoperability. *In 9th International Conference on Interoperability for Enterprise Systems and Applications (I-ESA 2018), March, 2018, Berlin, Germany.*

4. Krahn, T., Eichelberg, M., Müller, F., Gönül, S., Erturkmen, G. B. L., Sinaci, A. A., & Appelrath, H. J. (2014, September). Adverse drug event notification on a semantic interoperability framework. *In MIE (pp. 111-115).*

5. Sinaci, A. A., Erturkmen, G. B. L., Gönül, S., Cinar, H. A., & Kaya, A. (2013). Patient History Navigation with the Use of Common Data Elements. *In SWAT4LS.*

6. Yuksel, M., Gönül, S., Erturkmen, G. B. L., Sinaci, A. A., Depraetere, K., De Roo, J., & Bergvall, T. (2013). Demonstration of the SALUS Semantic Interoperability Framework for Case Series Characterization Studies. *In SWAT4LS.*

7. Gönül, S., & Sinaci, A. A. (2012, September). Semantic content management and integration with JCR/CMIS compliant content repositories. *In Proceedings of the 8th International Conference on Semantic Systems (pp. 181-184). ACM.*

8. Sinaci, A. A., & Gonul, S. (2012, May). Semantic Content Management with Apache Stanbol. *In Extended Semantic Web Conference (pp. 371-375). Springer, Berlin, Heidelberg.*

9. Tuncer, F., Dogac, A., Kabak, Y., POSTACI, Ş., Gönül, S., & Alpay, E. (2009, October). iSURF eDoCreator: e-Business Document Design and Customization Environment. *In proceedings of e-Challenges 2009 conference.*