

**Ioannis Chatzigiannakis
Boris De Ruyter
Irene Mavrommati (Eds.)**

LNCS 11912

Ambient Intelligence

**15th European Conference, Aml 2019
Rome, Italy, November 13–15, 2019
Proceedings**



Springer

Founding Editors

Gerhard Goos

Karlsruhe Institute of Technology, Karlsruhe, Germany

Juris Hartmanis

Cornell University, Ithaca, NY, USA

Editorial Board Members

Elisa Bertino

Purdue University, West Lafayette, IN, USA

Wen Gao

Peking University, Beijing, China

Bernhard Steffen

TU Dortmund University, Dortmund, Germany

Gerhard Woeginger 

RWTH Aachen, Aachen, Germany

Moti Yung

Columbia University, New York, NY, USA


More information about this series at <http://www.springer.com/series/7409>

Ioannis Chatzigiannakis ·
Boris De Ruyter · Irene Mavrommati (Eds.)


Ambient Intelligence

15th European Conference, AmI 2019
Rome, Italy, November 13–15, 2019
Proceedings

Editors

Ioannis Chatzigiannakis 
Sapienza University of Rome
Rome, Italy

Boris De Ruyter 
Philips Research
Eindhoven, The Netherlands

Irene Mavrommati 
Hellenic Open University
Patras, Greece

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-030-34254-8 ISBN 978-3-030-34255-5 (eBook)
<https://doi.org/10.1007/978-3-030-34255-5>

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

© Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume contains the papers presented at AmI 2019, the 15th edition of the European Conference on Ambient Intelligence, held in Rome, Italy during November 13–15, 2019. AmI 2019 is part of a series of annual conferences, which first took place in 2003 with the EUSAI event in Veldhoven, the Netherlands.

Ambient intelligence refers to normal working and living environments populated with embedded devices that can merge unobtrusively and in natural ways using information and intelligence hidden in the network connecting these devices (e.g. the Internet of Things). Such devices, each specialized in one or more capabilities, are intended to work together based on an infrastructure of intelligent systems, to provide a variety of services improving safety, security, and quality of life in ordinary living, traveling, and working environments.

This year’s theme was on “Data-driven Ambient Intelligence” that follows the vision of Calm Technology, where technology is useful but does not demand our full attention or interfere with our usual behavior and activities.

In response to the call for papers, a total 50 submission were received from researchers and practitioners from the fields of science, engineering, and design, all working towards the vision of Ambient Intelligence. Each submission was assigned to at least three Program Committee members, in some cases aided by subreviewers. Out of these, the committee decided to accept 20 full papers. The selection was made by the Program Committee based on originality, quality, and relevance to Ambient Intelligence. The accepted papers are included in this Vol. 11912 of Springer’s LNCS.

This year, AmI also solicited short papers of work in progress with substantial interest for the community. In total 11 short papers were accepted for publication, out of which 1 was retracted as the authors could not present it at the event. The selection of the short papers was made by the Program Committee.

Apart from the contributed talks, AmI 2019 included invited presentations by Fabio Paternò, Alex Gluhak, Salvatore Iaconesi, Oriana Persico, and John Pagonis.

Four special interest sessions and hands-on tutorials were held on November 13, 2019:

- Ambient Intelligence for Promoting Sustainable Behaviors
- Edge Machine Learning for Smart IoT Environments
- Ambient Intelligence in the Blockchain Era: How Decentralized Services Can Revolutionize the Traditional Client Server Paradigm in Automation and Ambient Intelligence
- Hands-on Tutorial on Embedded Artificial Intelligence

We wish to thank all authors who submitted their research to this conference, contributing to the high-quality program, the Program Committees for their scholarly effort, and all referees who assisted the Program Committees in the evaluation process.

We thank Springer for sponsoring the Best Paper Award, taking into account that the paper was selected by the Program Committee.

We are also grateful to Ugo Cinelli Sara Ciotti, Sabrina Giampaolletti, Venerino Filosa, Domenico Macari, and all the support staff of the Organizing Committee from the Sapienza University of Rome for helping us organize AmI 2019.

November 2019

Ioannis Chatzigiannakis
Boris De Ruyter
Irene Mavrommati

Organization

General Chairs

Ioannis Chatzigiannakis Sapienza University of Rome, Italy
Boris De Ruyter Philips Research, The Netherlands

Program Committee Chairs

Ioannis Chatzigiannakis Sapienza University of Rome, Italy
Boris De Ruyter Philips Research, The Netherlands
Irene Mavrommati Hellenic Open University, Greece

Workshop Chairs

Francesca Cuomo Sapienza University of Rome, Italy
Andreas Komninos University of Patras, Greece

Poster/Demos Chairs

Damianos Gavalas University of the Aegean, Greece
Dimitris Charitos National and Kapodistrian University of Athens,
Greece

PhD Forum Chairs

Federica Paganelli University of Pisa, Italy
Javed Vassilis Khan Technical University of Eindhoven, The Netherlands

Publicity Chairs

Monica Divitini NTNU, Norway
Carmelo Arditto University of Bari, Italy

Local Organization Chair

Andrea Vitaletti Sapienza University of Rome, Italy

Program Committee

Constantinos Marios Bournemouth University, UK
Angelopoulos
Carmelo Arditto University of Bari Aldo Moro, Italy

Dimitris Charitos	National and Kapodistrian University of Athens, Greece
Ioannis Chatzigiannakis	Sapienza University of Rome, Italy
Eleni Christopoulou	Ionian University, Greece
Francesca Cuomo	Sapienza University of Rome, Italy
Boris De Ruyter	Philips Research, The Netherlands
Monica Divitini	Norwegian University of Science and Technology, Norway
Abak A. Farsh	Norwegian University of Science and Technology, Norway
Damianos Gavalas	University of the Aegean, Greece
Dimitrios Gavalas	University of the Aegean, Greece
Sten Hanke	FH Joanneum GmbH, Austria
Josep Miquel Jornet	University at Buffalo, The State University of New York, USA
Dimitris Kalles	Hellenic Open University, Greece
Achilleas Kameas	Hellenic Open University, Greece
Vlasios Kasapakis	University of the Aegean, Greece
Javed Vassilis Khan	Eindhoven University of Technology, The Netherlands
Otilia Kocsis	University of Patras, Greece
Andreas Komninos	University of Patras, Greece
Aris Lalos	Industrial Systems Institute, Athena Research Center, Patras, Greece
Helen C. Leligou	Technological Educational Institute of Chalkis, Greece
Panos Markopoulos	Eindhoven University of Technology, The Netherlands
Irene Mavrommati	Hellenic Open University, School of Applied Arts, Greece
Georgios Mylonas	CTI Diophantus, Greece
Theofanis Orphanoudakis	Hellenic Open University, Greece
Federica Paganelli	University of Florence, Italy
Evi Papaioannou	University of Patras and CTI Diophantus, Greece
Dimitris Ringas	University of Ionion, Greece
Modestos Stavrakis	University of the Aegean, Greece
Georgios Styliaras	University of Patras, Greece
Panagiotis Trakadas	National and Kapodistrian University of Athens, Greece
Christos Tselios	University of Patras, Greece
Andrea Vitaletti	Sapienza University of Rome, Italy
Stelios Zerefos	Hellenic Open University, Greece

Additional Reviewers

Orestis Akrivopoulos
Dimitrios Amaxilatis
Gerasimos Arvanitis
George Birbilis
Domenico Garlisi
Nikos Kanakis
Panagiotis Kokkinos
Katja Kroeller
Andrea Lacava
Jacopo Maria Valtorta
Alessio Martino
Stavros Nousias
Chrysanthi Tziortzioti
Evangelos Vlachos

Contents

Ambient Lighting Atmospheres for Influencing Emotional Expressiveness and Cognitive Performance.	1
<i>Boris De Ruyter and Saskia Van Dantzig</i>	
Power Efficient Clock Synchronization in Bluetooth-Based Mesh Networks.	14
<i>Dmytro Makara, Vladyslav Tsybul'nyk, and Taras Kurnyts'kyi</i>	
Classifying Teachers' Self-reported Productivity, Stress and Indoor Environmental Quality Using Environmental Sensors.	27
<i>Johanna Kallio, Elena Vildjiounaite, Vesa Kyllönen, Jussi Ronkainen, Jani Koivusaari, Salla Muuraiskangas, Pauli Räsänen, Heidi Similä, and Kaisa Vehmas</i>	
User Requirements for the Design of Smart Homes: Dimensions and Goals	41
<i>Michaela R. Reisinger, Sebastian Prost, Johann Schrammel, and Peter Fröhlich</i>	
A Clustering Approach for Profiling LoRaWAN IoT Devices.	58
<i>Jacopo Maria Valtorta, Alessio Martino, Francesca Cuomo, and Domenico Garlisi</i>	
Experiences from Using Gamification and IoT-Based Educational Tools in High Schools Towards Energy Savings	75
<i>Federica Paganelli, Georgios Mylonas, Giovanni Cuffaro, and Ilaria Nesi</i>	
IL4IoT: Incremental Learning for Internet-of-Things Devices	92
<i>Yuanyuan Bao and Wai Chen</i>	
Enhanced Buying Experiences in Smart Cities: The SMARTBUY Approach	108
<i>Lorena Bourg, Thomas Chatzidimitris, Ioannis Chatzigiannakis, Damianos Gavalas, Kalliopi Giannakopoulou, Vlasios Kasapakis, Charalampos Konstantopoulos, Damianos Kypriadis, Grammati Pantziou, and Christos Zaroliagis</i>	
Action Recognition Using Local Visual Descriptors and Inertial Data	123
<i>Taha Alhersh, Samir Brahim Belhaouari, and Heiner Stuckenschmidt</i>	

CircuitsMaster: An Online End-User Development Environment for IoT Electronics 139
Ward Seetsen, Irene Mavrommati, and Vassilis-Javed Khan

Enabling Machine Learning Across Heterogeneous Sensor Networks with Graph Autoencoders. 153
Johan Medrano and Fuchun Joseph Lin

Development of an Acoustically Adaptive Modular System for Near Real-Time Clarity-Enhancement 170
Alexander Liu Cheng, Patricio Cruz, Nestor Llorca Vega, and Andrés Mena

Experiences from Using LoRa and IEEE 802.15.4 for IoT-Enabled Classrooms 186
Lidia Pocero, Stelios Tsampas, Georgios Mylonas, and Dimitrios Amaxilatis

Adaptive Service Selection for Enabling the Mobility of Autonomous Vehicles 203
Elif Eryilmaz, Manzoor Ahmed Khan, Frank Trollmann, and Sahin Albayrak

Discovering User Location Semantics Using Mobile Notification Handling Behaviour 219
Andreas Komminos, Ioulia Simou, Elton Frengkou, and John Garofalakis

Data-Driven Intrusion Detection for Ambient Intelligence. 235
Ioannis Chatzigiannakis, Luca Maiano, Panagiotis Trakadas, Aris Anagnostopoulos, Federico Bacci, Panagiotis Karkazis, Paul G. Spirakis, and Theodore Zahariadis

Spoken Language Identification Using ConvNets 252
Sarthak, Shikhar Shukla, and Govind Mittal


Indoor Air Quality and Wellbeing - Enabling Awareness and Sensitivity with Ambient IoT Displays 266
Andreas Seiderer, Ilhan Aslan, Chi Tai Dang, and Elisabeth André

ATHsENSE: An Experiment in Translating Urban Data to Multisensory Immersive Artistic Experiences in Public Space 283
Dimitris Charitos, Iouliani Theona, Penny Papageorgopoulou, Antonios Psaltis, Antonios Korosidis, Dimitris Delinikolas, Alexandros Drymonitis, Natalia Arsenopoulou, and Charalampos Rizopoulos

Characterization of Individual Mobility for Non-routine Scenarios from Crowd Sensing and Clustered Data	296
<i>Inês Cunha, João Simões, Ana Alves, Rui Gomes, and Anabela Ribeiro</i>	
A Flexible and Scalable Architecture for Human-Robot Interaction	311
<i>Diego Reforgiato Recupero, Danilo Dessì, and Emanuele Concas</i>	
Toward Supporting Food Journaling Using Air Quality Data Mining and a Social Robot	318
<i>Federica Gerina, Barbara Pes, Diego Reforgiato Recupero, and Daniele Riboni</i>	
Viewing Experience of Augmented Reality Objects as Ambient Media - A Comparison of Multimedia Devices	324
<i>Ilhan Aslan, Chi Tai Dang, Björn Petrak, Michael Dietz, Michael Filipenko, and Elisabeth André</i>	
Ranking Robot-Assisted Surgery Skills Using Kinematic Sensors	330
<i>Burçin Buket Oğul, Matthias Felix Gilgien, and Pınar Duygulu Şahin</i>	
uAQE: Urban Air Quality Evaluator	337
<i>Claudio Rossi, Alessandro Farasin, Giacomo Falcone, and Carlotta Castelluccio</i>	
Enhancing an Eco-Driving Gamification Platform Through Wearable and Vehicle Sensor Data Integration	344
<i>Christos Tselios, Stavros Nousias, Dimitris Bitzas, Dimitrios Amaxilatis, Orestis Akrivopoulos, Aris S. Lalos, Konstantinos Moustakas, and Ioannis Chatzigiannakis</i>	
A Distributed Multi-Agent System (MAS) Application For continuous and Integrated Big Data Processing	350
<i>Ariona Shashaj, Federico Mastroilli, Massimiliano Morrelli, Giacomo Pansini, Enrico Iannucci, and Massimiliano Polito</i>	
Human Activities Recognition Using Accelerometer and Gyroscope	357
<i>Anna Ferrari, Daniela Micucci, Marco Mobilio, and Paolo Napoletano</i>	
Towards Habit Recognition in Smart Homes for People with Dementia	363
<i>Gibson Chimamiwa, Marjan Alirezaie, Hadi Banaee, Uwe Köckemann, and Amy Loufi</i>	
Ambient Explanations: Ambient Intelligence and Explainable AI	370
<i>Jörg Cassens and Rebekah Wegener</i>	
Author Index	377



Ambient Lighting Atmospheres for Influencing Emotional Expressiveness and Cognitive Performance

Boris De Ruyter^(✉)  and Saskia Van Dantzig 

Philips Research, High Tech Campus, 5656 AE Eindhoven, The Netherlands
boris.de.ruyter@philips.com

Abstract. In this paper a study is reported for investigating the effects of a lighting atmosphere on *emotional expressiveness* and *cognitive processing*. An experimental lighting atmosphere was created for hospital consultation rooms to better support the shared decision making process of patient and clinician. The lighting atmosphere consists of two phases: (1) indirect, dimmed-warm light (supporting emotional expressiveness) and (2) direct, cold-bright light (supporting cognitive processing). The ambient lighting atmosphere was compared with a standard office lighting atmosphere involving 54 male participants. Participants took part in the experiment in pairs. During the first phase, they watched two emotion inducing film fragments and then discussed these fragments with each other. Under warm-dimmed lighting conditions significantly more emotions were expressed with less negative valence. During the second phase, participants performed two cognitive tasks. No statistical significant effects of lighting condition on both attention and concentration tasks were found.

The results showed that participants' emotions and anxiety level was influenced negatively by the film fragments in both conditions. During the discussion, participants in the intervention condition had significantly more eye contact and showed fewer negative expressions than participants in the control condition. On the cognitive tasks, there was no difference between the conditions, indicating that attention and concentration were not influenced by the intervention.

Keywords: Ambient lighting · Emotions · Cognitive performance

1 Introduction

As a vision, Ambient Intelligence (AmI) refers to environments that are sensitive and responsive to the presence of people [22]. Essential in AmI is its focus on the effects that applications of technologies, embedded in the environment, have on people. Examples of such effects include *feelings of immersiveness* when consuming content [22] and *being persuaded* for changing attitudes and behaviour [14]. The focus of this paper is on the effects lighting atmospheres have on people.

1.1 Lighting Atmospheres

Natural lighting is *complex in composition* (e.g. 5% Ultraviolet, 43% visible, 52% near-infrared), *cyclic* (e.g. circadian and circannual) and *variable* (e.g. intensity and color temperature). The advantage of artificial (i.e. electric) lighting is that different aspects can be manipulated to achieve specific effects of lighting.

Besides classical visual effects, light elicits non-visual brain responses, which influence *physiological* and *psychological* processes. A diversity of lighting effects on people have been reported in literature. These range from immediate effects such as enhanced concentration by children in a classroom [24] and mood regulation [15] to longer term effects such as the regulation of sleep in hospitalized patients [12].

Studies have found relationships between lighting and *stress* [31], *depression* [27], *emotions* [9] and *circadian rhythm* [7]. In turn, these relationships influence different outcomes such as *length-of-stay* in hospitals [1] and *burnout* in working environments [3]. A conceptual overview of these relationships and their outcomes is presented in Fig. 1 based on findings from literature [1,3,5,7-9,11,13,20,27,29-32,34].

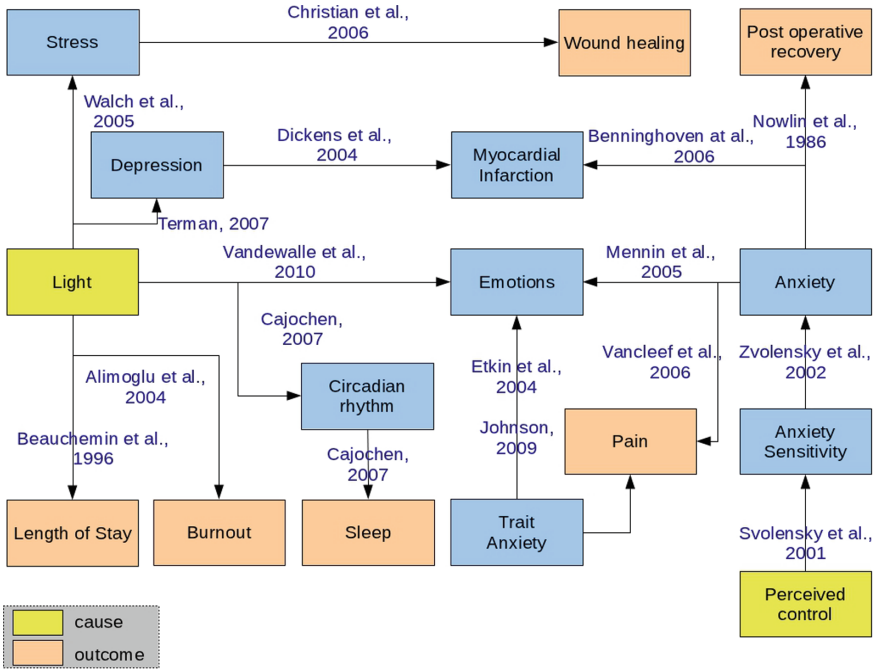


Fig. 1. A conceptual overview of lighting effects and outcomes documented in literature (mediators are marked in blue boxes) (Color figure online)

For the context of the study reported in this article, the focus is on the influence of lighting on processing of and emotional coping with medical diagnostic

information by patients recently diagnosed with prostate cancer. This type of cancer can be treated by a multitude of interventions. Since the most appropriate intervention is selected on individual basis, an active role of the patient is required in a shared decision process.

1.2 Shared Decision Making

Shared decision making is an emerging approach in healthcare, in which the clinician provides the patient with the complete clinical evidence related to the medical condition, the available treatment options, and the benefits and risk factors of each option. The patient shares personal preferences on the trade-off between the different benefits and risk factors. Together, the clinician and patient make an informed decision on a treatment pathway. Shared decision making is the recommended approach for prostate cancer, and its becoming increasingly common in medical practice [4]. In many cases of localized prostate cancer, multiple treatment alternatives are possible. Each of these alternatives has its own advantages and disadvantages, which may be weighed differently by different patients. Therefore, the patients personal situation and preferences should be taken into account to select the optimal solution on a case by case basis.

The decision making process generally consist of the following steps:

1. **Consultation and tests.** During the first consultation with the urologist, the patients complaints, health situation and risks are discussed. A physical exam and biopsy are performed or scheduled and the patient is asked to come back in a week for the diagnosis.
2. **Diagnosis.** After one week, the urologist shares the test results with the patient, explains the diagnosis and describes the possible treatment options.
3. **Reflection.** The patient (and partner) retreat to a reflection room, where they are joined by a nurse, who helps them to cope with their emotions and who ensures that the patient has understood the diagnosis and the different treatment options.
4. **Thinking period.** The patient then has a period of several weeks to think over his options. Patients often search the internet, talk to family and friends, and they may consult the nurse and other specialists in the hospital or take a 2nd opinion. In the meantime, the urologist meets with the radiologist and oncologist to discuss the treatment options.
5. **Decision.** During the next consultation, the urologist asks the patient about treatment preferences and shares the conclusions from the experts meeting. Together, the patient and the urologist decide which treatment path will be followed.

While it is desirable to adopt the above described shared decision making process, due to the emotional distress caused by receiving such diagnosis patients are not always in an optimal condition for participating in such process. Building on established influences of lighting on cognitive processing and emotional

experiences, an ambient lighting atmosphere is developed which could facilitate the shared decision making process. This ambient lighting atmosphere could be deployed in different hospital spaces, such as waiting rooms, consultation rooms (where patients meet with the urologist) and reflection rooms (where patients meet with the nurse after receiving the diagnosis). The different lighting atmospheres are depicted in Fig. 2.



Fig. 2. Different lighting atmospheres

2 Study

The study investigates the potential positive effects of a lighting atmosphere on *emotional expressiveness* (by enhancing relaxation) and *cognitive information processing* (by enhancing alertness).

The study protocol was reviewed and approved by an ethics committee and followed the Declaration of Helsinki (World Medical Association, 1964). Participant consent was obtained prior to the start of the study.

2.1 Lighting Atmosphere

The experimental lighting atmosphere implements two sequential lighting settings: **dimmed-warm** and **cold-bright** light (see Table 1). Based on the following literature, it is expected that dimmed-warm light will influence emotional experiences while cold-bright light will influence information processing:

- Low illuminance environments (150 Lux) induce a global processing style (the tendency to focus on the big picture). In high illuminance environments (1500 Lux) a local processing style (the tendency to attend to details) is induced [26]
- Under low levels of lighting more positive affect, feelings of relaxation and intimacy are reported compared to high levels of lighting [2,10]
- Dimmed warm light (2700 K) leads to the impression of a relaxing and intimate atmosphere, while bright cool light (5000 K) is associated with comfort and spaciousness [19]

Table 1. Lighting settings implemented in the study

	Intervention condition		Control condition
	Relaxing atmosphere	Focusing atmosphere	
Description	Indirect, warm light	Direct, cool light	standard office lighting
Color temp.	2700 K	5000 K	
Illuminance (horizontal)	150 Lux	1500 Lux	
Direction	Wall mounted, up-light	Ceiling mounted, downwards	

2.2 Participants

Fifty-four Dutch speaking men were recruited by an external agency for participation in the study. Participants were between 46 and 64 years old and had an average age of 55.8 years ($SD = 5.9$). In return for their participation, participants received a gift voucher of 15 Euros and travel expenses. Colorblind subjects, users of sleep medication, shift-workers or subjects suffering from an eye disease were excluded from the study.

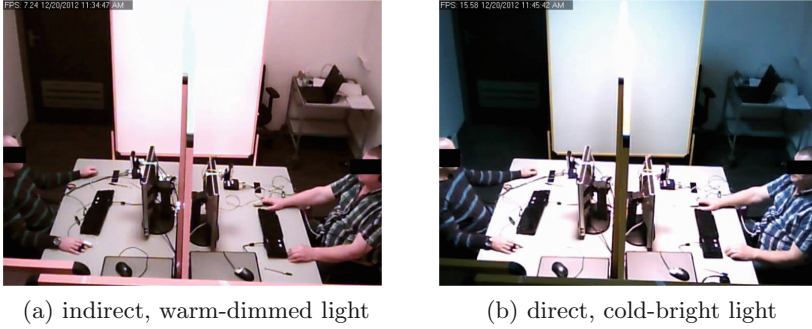
2.3 Study Design

The experiment consisted of two phases; the first phase focusing on the elicitation and expression of emotion, and the second phase on cognitive processing (both phases are typically part of the *reflection* step during the decision making process as described above). In the control condition, normal office lighting was presented throughout the whole experiment. In the intervention condition a dynamic lighting atmosphere consisting of two phases was used. During the first phase, a relaxing atmosphere was created through indirect, dimmed-warm light. During the second phase, a focusing atmosphere was created through direct, cold-bright light. The intervention and control condition are implemented following a between-subjects design.

The hypothesis to be tested are:

- emotional expressiveness will be different under dimmed-warm lighting conditions compared to under neutral office lighting
- cognitive processing will be different under cold-bright light conditions compared to neutral office lighting (Fig. 3).

During the first phase (i.e. under dimmed-warm light) of the experiment, *feelings* and *anxiety levels* were measured before and after watching emotion inducing film fragments. In the second phase (i.e. cold-bright light) of the experiment, cognitive tests for measuring *attention* and *concentration* were administered (see Table 2). The instruments used in the study are discussed together with the study procedure in the next section.

**Fig. 3.** Intervention settings**Table 2.** Data acquisition during the study

Phase	Task	Goal
1	3-item Feeling Scale	Measure current emotional state
	20-item trait anxiety scale	Measure current anxiety level
	Watch video fragments	Induce emotions
	Discuss video fragments	Measure emotional expression
after inducing emotions (T_2)	3-item Feeling Scale	Measure current emotional state
	20-item trait anxiety scale	Measure current anxiety level
2	N-back	Measure attention
	D-2	Measure concentration level

2.4 Procedure

The experiment took place in a controlled laboratory room and lasted for approximately 50 min. Participants took part in the experiment in pairs. Each pair was randomly assigned to either the intervention condition or the control condition. In total, there were 24 participants in the intervention condition and 30 participants in the control condition.

Upon entering the laboratory room, participants were seated behind two tables facing each other (see Fig. 4). On each table there was a computer monitor, input devices (keyboard & mouse) and a camera. A screen was placed between the two tables, preventing the participants from seeing each other directly. However, at certain moments during the study, they could view and communicate with each other via a camera mounted on a monitor.

When the participants were seated, the project leader briefly explained the procedure of the study, clarified the informed consent form and asked participants to sign the form. Then, participants put on headphones and started with the actual experiment.

First, participants responded to some questions with regard to their current emotional state and level of anxiety. Emotional state was assessed by the Feeling Scale, which contains three items relating to *overall valence* (very bad \leftrightarrow very good), *energetic arousal* (tired \leftrightarrow energetic) and *tense arousal* (stressed

↔ relaxed) [33]. Responses are given on a nine-point Likert scale. The anxiety level was assessed by 20 items of the State-Trait Anxiety Inventory (STAI), assessing the participants state level of anxiety, answered on a four-point Likert scale [21, 25]. Next, participants watched two emotion inducing film fragments (selected to induce negative emotions) on their computer monitors [23].

After watching two emotion inducing film fragments (for a validation of emotion induction using video fragments see [16]), participants removed their headphones and discussed the fragments with the other member of the pair, for a duration of three minutes. This discussion was intended to elicit emotional expression. During the discussion, the participants faces were captured by a camera mounted on the monitor, such that they could see each other on their computer monitors. In addition, the video feed from both cameras was used for an automated assessment of the emotional expressiveness based on facial expressions. For this assessment the FaceReader 3.0 tool from Noldus B.V. was used. This tool provides a classification of facial expressions in terms of direction and strength of emotional expressiveness [17, 28] and can distinguish six basic emotions (*happy, angry, sad, surprised, scared, disgusted*) plus neutral, with an accuracy of 88%.

Following the discussion, participants completed again the *feeling* and *anxiety* questionnaire after which the second phase of the experiment started, addressing cognitive performance. While in the control condition the lighting remained constant, in the intervention condition there was a sudden switch to cool bright light, which was expected to enhance attention and concentration. During the second phase participants completed (individually) two cognitive tasks: (i) the N-back task, a computerized attention task and (ii) the D-2, a paper and pen task measuring attention [6]. After completion of the tasks, the test leader briefed the participants about the objectives of the study, thanked them for their participation and handed them a gift voucher.

3 Results

3.1 Manipulation Check

Before going into the statistical analysis of the study results, some manipulation and quality checks of the obtained data are needed.

Quality of the Facial Expression Recordings. Starting from the raw data (as provided from the FaceReader analysis per frame of the recorded video) a quality index (based on the ratio of available frames and number of frames for which the FaceReader was able to obtain an emotional expressiveness and valence value) is calculated. Although the quality of a video recording for analysis with the FaceReader software can be influenced by many factors (e.g. gaze direction, occlusion of parts of the face by hands, facial features such as beard) it is documented that lighting conditions can significantly influence the quality of the analysis. Hence, it is important to verify that the quality of the video

recordings is comparable for both experimental (and thus lighting) conditions. A statistical analysis (using a t -test) of the significance difference between the quality index of both conditions indicates that there is no significant difference between the quality index for both lighting conditions ($t = 0.4, p = 0.692$).

Inducing Emotions with Film Fragments. The emotional state was assessed with the Feeling Scale (measuring *overall valence*, *energetic arousal* and *tense arousal*) before (T1) and after (T2) viewing the emotion inducing film fragments (see Table 2).

On the *overall valence* item, participants responded significantly lower (using a repeated measures ANOVA test) at T2 than at T1, $F(1, 51) = 12.9, p < 0.001$. There was no main effect of condition, $F(1, 51) = .76, p = 0.40$, nor an interaction between time (before and after video fragments) and condition, $F(1, 51) = 2.9, p = 0.10$.

On the *energetic arousal* item, there was no significant difference between T1 and T2, $F(1, 51) = 2.9, p = 0.09$. There was also no effect of condition, $F(1, 51) = 2.7, p = 0.11$, or an interaction between time and condition $F(1, 51) = 0.42, p = 0.52$.

On the *tense arousal* item, there was a significant effect of time, $F(1, 51) = 7.4, p = 0.01$. Participants reported feeling more stressed at T2 than at T1. There was no significant main effect of condition, $F(1, 51) = 2.0, p = 0.16$, nor a significant interaction between time and condition, $F(1, 51) = .05, p = 0.83$.

The results of the Feeling Scale suggests that participants emotions were indeed influenced by watching the film fragments: participants felt slightly worse and more tense after watching the fragments than at the start of the experiment.

3.2 Emotional Expression

Emotional expression was assessed (from recordings during the discussion between participants after watching the film fragments) with the Noldus FaceReader, which classifies facial expressions into six basic emotions or a neutral expression. Due to camera and software issues, some sessions were not recorded, resulting in 23 recordings in the intervention condition, and 22 in the control condition.

The FaceReader software provides per frame in the analysis an index expressing the probability or index of a **neutral** and an **emotional** facial expression. A statistical analysis between the neutral index for both groups indicated that there is no significant difference in the amount of **neutral expressiveness** for both groups ($t = -1.7, p = 0.098$). A statistical analysis between the **emotional expressiveness** index for both groups indicates that although more emotions were expressed in the control condition compared to the intervention condition, this difference is not statistically significant ($t = 1.98, p = 0.055$). Finally, the FaceReader software provides an index for the **valence** of the emotional expressiveness per frame included in the analysis. A statistical analysis indicates that there is no significant difference between for positive valence index ($t = -0.34,$

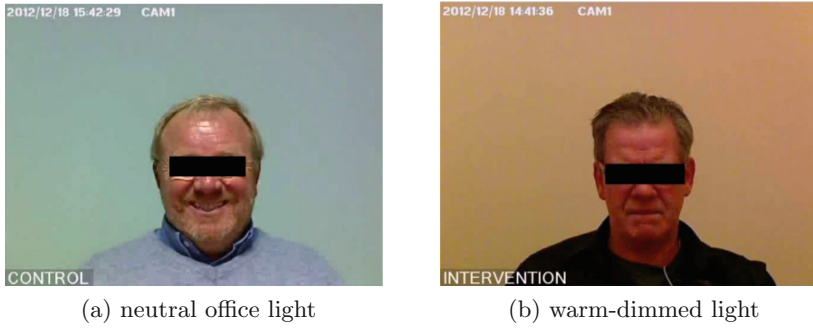


Fig. 4. Capturing of facial expression

$p = 0.735$) while there is a significant difference for the negative valence index ($t = -2.54$, $p = 0.015$) between the control and intervention group.

During the study it was noted that some participants avoided eye contact (based on the gaze direction information provided by the FaceReader). Interestingly, the ratio of participants avoiding eye contact differed between conditions. In the intervention condition, 1 out of 23 participants (4%) avoided eye contact, whereas 7 out of 22 participants (32%) in the control condition avoided eye contact. This difference in ratio was significant, $\chi^2 = 5.8$, $p = 0.016$.

3.3 Anxiety

Anxiety was assessed with the State Trait Anxiety Questionnaire (STAI), which consists of 40 items (20 items measuring trait anxiety and 20 items measuring state anxiety). Participants answered the 20 state anxiety items twice, before (T1) and after (T2) viewing the emotion inducing videos. The STAI was answered by 25 participants in the Intervention condition and 28 participants in the Control condition. The data were analysed with a 2×2 repeated measures Analysis of Variance (ANOVA), with time (T1 vs. T2) as within-subjects factor and condition (intervention vs. control) as between-subjects factor.

There was a significant main effect of time, $F(1, 51) = 6.4$, $p = 0.015$. Participants reported higher anxiety levels after than before viewing the emotion inducing video fragments. There was no main effect of condition, $F(1, 51) = 0.4$, $p = 0.54$, nor an interaction between time and condition, $F(1, 51) = 0.05$, $p = 0.82$. The results indicate that participants felt slightly more anxious after watching the film fragments, irrespective of experimental condition.

3.4 Cognitive Performance

The hypothesis with respect to the second phase was that cognitive processing should be enhanced when participants are exposed to bright cold light. In other words, participants should perform better on the cognitive tasks (concentration and attention) in the intervention condition compared to the control condition.

Concentration. Concentration was measured with the N-back task. The data from 11 participants were removed from the analysis because they had an accuracy level below 40%. The scores from 18 participants in the intervention condition and 24 participants in the control condition remained and were analysed statistically with an independent samples *t*-test. There was no statistical significant difference between the experimental conditions with respect to accuracy or reaction time (all *t*-scores lower than 1).

Attention. Attention was measured with the D2 test. The data from two participants were removed from the analysis (one failed to perform the task correctly and one did not complete the task). Next, based on Tukeys criterion, two outlying scores were removed. The scores from 28 participants in the Control condition and 23 participants in the Intervention condition remained and were analysed statistically, using an independent samples *t*-test. There was no significant difference between the two experimental conditions on the total number of processed items, the error percentage, or the concentration score (all *t*-scores lower than 1.0). From these results, we can conclude that the intervention did not enhance attention.

4 Discussion

For a shared decision making process to be successful, patients should be actively involved in the process. However, often patients are not in the most suitable emotional and cognitive state to effectively participate in the process. Current practice does not give patients the time and space to express their emotions upon receiving the bad news. Instead, they have to process information about their diagnosis, and discuss the possible treatment options with their clinician. Because patients are typically in shock after receiving the bad news, they have difficulties assimilating and remembering the provided information. As a result, they often ask repetitive questions, putting an additional burden to clinicians.

To address these issues, we created and tested a dynamic lighting atmosphere for consultation rooms that should enhance (a) the expression of emotions, and (b) the processing of information. This method builds on several well documented effects of lighting on the expression of emotions and the processing of information. Our hypotheses were tested in a laboratory study, in which participants discussed emotion inducing film fragments and performed cognitive tasks, while being exposed to either the dynamic lighting atmosphere (intervention condition) or standard office lighting (control condition).

The results of this experiment show that participants emotion and anxiety levels were influenced by the film fragments in both conditions. After viewing the film fragments (compared to before viewing the fragments) there was significantly more **anxiety**, **lower valence** and higher **tense arousal**.

During the discussion, participants in the intervention condition had statistically significantly **more eye contact** and **fewer negative emotional expressions** than participants in the control condition. While not statistically signifi-

cant ($p = 0.055$), there were more emotions expressed under warm-dimmed light compared to the standard office lighting condition.

Although the alerting effects of bright light are well acknowledged in literature [7,18], the current study found no difference between the conditions, indicating that attention and concentration were not influenced by the intervention. It could be argued that test participants were already in a higher level of alertness due to the emotions induced and expressed in the first part of the experiment. A possible ceiling effect could explain the observation that the bright light condition during the second part of the experiment did not increase alertness.

5 Conclusion

In general, it can be concluded that the warm-dimmed lighting atmosphere significantly influenced positively the expression of emotions (even though the film fragments were selected to induce negative emotions). Additionally, the same film fragments induced significantly less negative valence in the facial expressions (and were thus potentially perceived as less negative) under the warm-dimmed lighting atmosphere. Additionally, the same film fragments induced significantly less negative valence in the facial expressions (and were thus potentially perceived as less negative) under the warm-dimmed lighting atmosphere. In the present study, no statistically significant effects were found on cognitive processing (attention and concentration).

There are some limitations that apply to this study. It is likely that the induced levels of emotion and anxiety in a laboratory context are less intense than the experienced levels of emotion and anxiety in actual patients during the shared decision making process. Thus, before bringing the proposed ambient lighting atmosphere into the actual hospital context, we should investigate if the ambient lighting atmosphere does have a beneficial effect for the targeted patients. Studies with actual patients will have to provide evidence that such envisaged effect is of significance (i.e. having a large enough impact to positively influence the shared decision making process) for real-life settings such as those in a hospital.

The aim of the intervention was to create an atmosphere in which people feel comfortable to share their emotions. In the warm-dimmed lighting atmosphere, participants had more eye-contact and showed fewer negative expressions than in the control condition. This raises the question whether participants in this condition actually perceived less negative emotions, or whether they withheld their emotions more. To address this question, further research is needed.

There are different ways to implement the ambient lighting atmosphere in hospital rooms. In its simplest embodiment, the appropriate lighting setting is selected by the clinician (e.g., through a remote control). In a more intelligent embodiment, the system could contain sensors to monitor the patient's emotional response and/or conversation topic. It could respond intelligently to the information gathered in the room. Obviously, in such a solution, the patient's privacy is of utmost importance and should be guarded strictly.

References

1. Alimoglu, M., Donmez, L.: Daylight exposure and the other predictors of burnout among nurses in a University Hospital. *Int. J. Nurs. Stud.* **42**(5), 549–555 (2005)
2. Baron, R.: Lighting as a source of positive affect (1990)
3. Beauchemin, K., Hays, P.: Sunny hospital rooms expedite recovery from severe and refractory depressions. *J. Affect. Disord.* **40**, 49–51 (1996)
4. Bekelman, J.E., et al.: Clinically localized prostate cancer: ASCO clinical practice guideline endorsement of an American urological association/American society for radiation oncology/society of urologic oncology guideline. *J. Clin. Oncol.* **36**(32), 3251–3258 (2018). <https://doi.org/10.1200/JCO.18.00606>
5. Benninghoven, D., Kaduk, A., Wiegand, U., Specht, T., Kunzendorf, S.: Influence of anxiety on the course of heart disease after acute myocardial infarction risk factor or protective function? *Psychother. Psychosom.* **75**, 56–61 (2006)
6. Brickenkamp, R., Zillmer, E.: Test D2: Concentration-Endurance Test. CJ Hogrefe Publishers, Gottingen (1998)
7. Cajochen, C.: Alerting effects of light. *Sleep Med. Rev.* **11**(6), 453–464 (2007)
8. Christian, L., et al.: Stress and Wound Healing. *Neuroimmunomodulation* **13**(5–6), 337–346 (2006)
9. Dickens, C., et al.: The risk factors for depression in first myocardial infarction patients, vol. 34. *Psychol Med* (2004)
10. Durak, A., Olguntrk, N., Yener, C., Gven, D., Grnar, Y.: Impact of lighting arrangements and illuminances on different impressions of a room. *Build. Environ.* **42**(10), 3476–3482 (2007)
11. Etkin, A., et al.: Individual differences in trait anxiety predict the response of the basolateral amygdala to unconsciously processed fearful faces. *Neuron* **44**(6), 1043–1055 (2004)
12. Gimenez, M., et al.: Light and sleep within hospital settings, sleep-wake research in the Netherlands. *Ann. Proc. Dutch Soc. Sleep-Wake Res. (NSWO)* **22**, 56–59 (2011)
13. Johnson, D.: Emotional attention set-shifting and its relationship to anxiety and emotion regulation. *Emotion* **9**(5), 681 (2009)
14. Kaptein, M.C., Markopoulos, P., de Ruyter, B., Aarts, E.: Persuasion in ambient intelligence. *J. Ambient Intell. Humaniz. Comput.* **1**(1), 43–56 (2010). <https://doi.org/10.1007/s12652-009-0005-3>
15. Kuijsters, A., Redi, J., de Ruyter, B., Heynderickx, I.: Lighting to make you feel better: improving the mood of elderly people with affective ambiances. *PLoS ONE* **10**, e0132732 (2015)
16. Kuijsters, A., Redi, J., de Ruyter, B., Heynderickx, I.: Inducing sadness and anxiousness through visual media: measurement techniques and persistence. *Front. Psychol.* **7**, 1141 (2016). <https://doi.org/10.3389/fpsyg.2016.01141>. <https://www.frontiersin.org/article/10.3389/fpsyg.2016.01141>
17. van Kuilenburg, H., Wiering, M., den Uyl, M.: A model based method for automatic facial expression recognition. In: Gama, J., Camacho, R., Brazdil, P.B., Jorge, A.M., Torgo, L. (eds.) *ECML 2005. LNCS (LNAI)*, vol. 3720, pp. 194–205. Springer, Heidelberg (2005). https://doi.org/10.1007/11564096_22
18. Lehl, S., et al.: Blue light improves cognitive performance. *J. Neural Transm.* **114**(4), 457–460 (2007)
19. Manav, B.: An experimental study on the appraisal of the visual environment at offices in relation to colour temperature and illuminance. *Build. Environ.* **42**(2), 979–983 (2007)

20. Mennin, D., Heimberg, R., Turk, C., Fresco, D.: Preliminary evidence for an emotion dysregulation model of generalized anxiety disorder. *Behav. Res. Ther.* **43**(10), 1281–1310 (2005)
21. van der Ploeg, H.: The development and validation of the Dutch State-Trait Anxiety Inventory: ‘De Zelf-Beoordelings Vragenlijst’ (1981)
22. de Ruyter, B., Aarts, E.: Ambient intelligence: visualizing the future. In: *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI 2004*, pp. 203–208. ACM, New York (2004). <https://doi.org/10.1145/989863.989897>
23. Schaefer, A., Nils, F., Sanchez, X., Philippot, P.: Assessing the effectiveness of a large database of emotion-eliciting films: a new tool for emotion researchers. *Cogn. Emot.* **24**(7), 1153–1172 (2010)
24. Slegers, P., Moolenaar, N., Galetzka, M., Pruyn, A., Sarroukh, B., Zande, B.: Lighting affects students concentration positively: findings from three Dutch studies. *Lighting Res. Technol.* (2012)
25. Spielberger, C.D., Gorsuch, R.L., Lushene, R., Vagg, P.R., Jacobs, G.A.: *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press, Palo Alto (1983)
26. Steidle, A., Werth, L., Hanke, E.V.: You cant see much in the dark. *Soc. Psychol.* **42**(3), 174–184 (2011)
27. Terman, M.: Evolving applications of light therapy. *Sleep Med. Rev.* **11**(6), 497–507 (2007)
28. Terzis, V., Moridis, C.N., Economides, A.A.: Measuring instant emotions during a self-assessment test: the use of facereader. In: *Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research, MB 2010*, pp. 18:1–18:4 (2010)
29. Vancleef, L., Peters, M.: Pain catastrophizing, but not injury/illness sensitivity or anxiety sensitivity, enhances attentional interference by pain. *J. Pain* **7**, 23–30 (2006)
30. Vandewalle, G., et al.: Spectral quality of light modulates emotional brain responses in humans. In: *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp. 19549–19554 (2010)
31. Walch, J., Rabin, B., Day, R., Williams, J., Choi, K., Kang, J.: The effect of sunlight on postoperative analgesic medication use: a prospective study of patients undergoing spinal surgery. *Psychosom. Med.* **67**, 156–163 (2005). (1534–7796. Comparative Study PT - Journal Article PT - Research Support, Non-U.S. Govt)
32. Wells, J., Howard, G., Nowlin, W., Vargas, M.: Presurgical anxiety and postsurgical pain and adjustment: effects of a stress inoculation procedure. *J. Consult. Clin. Psychol.* **54**(6), 831 (1986)
33. Wilhelm, P., Schoebi, D.: Assessing mood in daily life. *Eur. J. Psychol. Assess.* **23**(4), 258–267 (2007)
34. Zvolensky, M., Feldner, M., Eifert, G., Stewart, S.: Evaluating differential predictions of anxiety-related reactivity during repeated 20% carbon dioxide-enriched air challenge. *Cogn. Emot.* **15**, 767–786 (2001)



Power Efficient Clock Synchronization in Bluetooth-Based Mesh Networks

Dmytro Makara¹, Vladyslav Tsybul'nyk¹, and Taras Kurnyts'kyi

R&D Departement, SoftServe Inc., V.Velykogo 52, Lviv, Ukraine
{dmakara, vtsyb, tkurnyt}@softserveinc.com

Abstract. Power efficiency is a hot topic nowadays due to dramatically increasing number of devices (nodes) that are difficult to charge, so they need to harvest energy from the environment itself. Since the amount of harvested energy is very limited, it is impossible for the nodes to even “listen” continuously for the packets from neighbors. For the proper scheduling of communication, clock synchronization between the nodes is required. This paper describes the clock synchronization design for WSN in case the continuous listening is not possible. Practical power consumption measurements based on Nordic nRF52840 chip are provided. Parameters selection to minimize the clock synchronization time and power consumption are investigated and example calculations of Lightricity solar panel are provided for reference. Using only 140 μW from this panel, designed clock synchronization allows 2 nodes to be synchronized within 180s time interval with typical 10 ms error.

Keywords: Clock synchronization · Power efficiency · Mesh networks · Bluetooth Low Energy

1 Introduction

Modern communication technologies make it possible to substantially reduce power consumption in comparison with the conventional approaches. This contributed to the emergence of the ultra-low-power networks which nodes can operate for up to a year on a coin battery.

On the other hand, with the ever-growing use of renewable energies, especially solar energy, there is a demand for the further development of completely battery-less networks operating solely from energy harvesting.

Time synchronization between the nodes is an important part of the proper functioning of any network, both wired and wireless. Indeed, common sense of time, especially in wireless mesh networks, is rather a basic requirement than a “nice-to-have” feature.

Apart from this, time synchronization can serve multiple useful purposes in wireless mesh networks. Usually, when a sensor network is deployed the nodes relative locations are unknown. With the implemented proper time synchronization, the distances between the nodes can be established by send/receive message

timestamps. Moreover, the timestamps can also be used for node's mobility and speed detection.

Time synchronization is also important because of the imperfection of hardware clocks. As time elapses, the clocks drift gradually from each other thus requiring periodical resynchronization.

However, the most important role the time synchronization plays in the low-power mesh networks, where the nodes fall asleep to save the energy. Nodes in these networks are inactive for most of the time and wake up only for a very short period to exchange messages. Correct time synchronization becomes extremely important in order to assure the synchronous sending/receiving time slots. On the other hand, this imposes extra constraints on time synchronization scheme making usage of the ready time sync solutions practically impossible.

In this paper we propose the time synchronization design for power-efficient BLE sensor mesh networks. Existing time and clock synchronization protocols are tailored for time synchronization accuracy and are not suitable for networks with the strict power consumption limitations, e.g. based on battery-less nodes. Proposed lightweight clock synchronization desing minimizes as much as possible the radio activity, that consumes the majority of energy. Practical measurements are provided and serve as basics for parameter selection recommendations.

2 Time Synchronization in WMN

Nowadays, Bluetooth is very popular for wireless sensor nodes. It is widely adapted for the low-range wireless communication in consumer devices and, as an open standard, it has great support of the industry and becomes ubiquitous.

In 2017, the Bluetooth SIG released the Bluetooth mesh standard. There are several BLE mesh SDKs developed by different vendors. For example, Nordic Semiconductors, Silicon Labs, and some others developed their SDKs based on Bluetooth mesh standard. According to those, nodes with different roles are introduced. Among those, there are also low power nodes sleeping most of the time with the radio being turned off. Other nodes, however, have radio switched on permanently which is unacceptable for the ultra-low power networks. To reduce power consumption, the nodes should wake up synchronously to exchange data, thus time synchronization within network is extremely important.

2.1 Overview of Time Synchronization Protocols

Network Time Protocol (NTP) and Precise Time Protocol (PTP) are well-known standard time synchronization network protocols. However, they are not applicable to wireless networks due to power consumption issues which are critical for the latter. Among the diversity of the wireless media and technologies (Wi-Fi, Bluetooth, Bluetooth Low Energy, ZigBee, etc.) and various time synchronization requirements, current standard time sync protocols for wireless networks do not exist. Only de-facto standards exist. Moreover, in the specific cases custom

protocols are devised [5]. For example, direct porting to BLE of time sync protocol has proved efficient for Bluetooth, but it may not be applicable due to the specific power consumption conditions in BLE networks. Choice of the proper time sync protocol is to the large extent dictated by the required time sync accuracy. In general, the looser time accuracy requirements are the more light-weight protocol (in terms of complexity and power consumption) is applicable [11–13].

One of the most important challenges in the context of WSN time synchronization is the internal clock drift. It is caused by the fact that crystal oscillators (the source for most clocks) have slightly different frequencies from device to device, even if they are configured to be identical. As time elapses, these frequencies drift from each other thus producing time offsets, that is, difference in time readings between the clocks.

The other important WSN time synchronization problem is non-determinism. It means that the time needed to deliver message from one node to another in the network varies depending on the specific conditions. Namely, traffic level, processor load, interrupts, physical media accessibility, etc. cause the total message traveling time to vary within a wide range. This makes the time synchronization problem more difficult to solve. One of the general solutions to this problem is generating timestamps at OSI level as close to physical level as possible.

2.2 Time Synchronization Protocols for Wireless Networks

In conventional wired networks, where the nodes are permanently active and there are no power consumption constraints, the time synchronization problem is mainly a matter of a time synchronization algorithm. In wireless networks with the battery-powered nodes this problem gets more complex due to the issue of power consumption. With WSN over BLE, the problem is getting even more complex since nodes are not permanently active and most of the time they sleep to conserve energy and cannot exchange messages. As for wireless networks in general and WSN, in particular, there are no official time synchronization standards. Instead, a couple of listed below de-facto standards exist [14]:

- Reference Broadcast Synchronization (RBS)
- Time Sync Protocol for Sensor Networks (TPSN)
- Flooding Time Synchronization Protocol (FTSP)
- Tiny-Sync
- Mini-Sync

Overview of RBS and TPSN clock synchronization methods is provided in [3].

Reference Broadcast Synchronization. This is a rather simple time sync approach which is based on the receiver-receiver pattern rather than on the sender-receiver. In such a way it eliminates send and access times as sources of variability in message delivery duration. Briefly, the protocol works in the following way: sender periodically broadcasts sync message to other nodes; the

nodes timestamp message arrival according to their individual clocks; afterwards, they exchange another sync message with each other. Therefore, each node knows other nodes' offsets from its own clock. To the best of our knowledge, no open source implementations of RBS are available. Both TPSN and FTSP use ideas from RBS.

Time Sync Protocol for Sensor Networks. TPSN is a sender-receiver synchronization protocol. The network structure has a tree topology. TPSN provides network-wide time synchronization in a sensor network. The TPSN algorithm consists of two stages:

- level discovery stage when the hierarchical network structure is discovered;
- synchronization phase when the pair-wise synchronization is carried out.

TPSN is designed for multi-hop networks, while RBS performs best on the single-hop networks. TPSN uses MAC layer timestamping. During the synchronization phase, the two-way message exchange between a pair of nodes forms the main part of the algorithm. Basing on this, the offset time between nodes' clocks is calculated. Unlike RBS, TPSN implementation can be found in open sources. More detailed description of TPSN can be found in [2].

Flooding Time Synchronization Protocol. This is another sender-receiver protocol. It is similar to TPSN in that it also has the root node selected and all nodes are synchronized to the root. However, unlike TPSN which has a tree topology, FTSP has a mesh topology. FTSP is best suited for the large multi-hop networks. FTSP uses MAC layer timestamping like TPSN. However, this requires a custom-made stack implementation to enable timestamping at this level. FTSP has several advantages over TPSN. Although TPSN provides a protocol for a multi-hop network, it does not handle the topology changes well. TPSN would have to re-initiate the level discovery phase if the root node or the topology was changed. This would require more network traffic and would create additional overhead.

With FTSP, the flooding aids in handling dynamic topology changes. The protocol specifies that the root node is periodically re-elected. Like TPSN, FTSP also provides MAC layer time stamping which greatly increases the precision and reduces jitter.

FTSP is an available feature in TinyOS [4] which is an open source, event-driven, and modular OS designed to be used with the sensor networks.

Tiny-Sync and Mini-Sync. Both the Tiny-Sync and Mini-Sync algorithms utilize the traditional synchronization method involving the two-way messaging between a pair of nodes for the interpretation of the relative drift and relative offset between the clocks. Though they are not as popular as the previously described methods, they are mentioned here because these algorithms can be useful for fast implementation and benchmarking.

2.3 Time Synchronization in Low-Power Mesh Networks

The choice of the time sync protocol for WMN over BLE is dictated by the specific application requirements. It is usually a matter of trade-off between the power consumption and timestamping accuracy. All the protocols described above are tailored for the time sync accuracy and are not good candidates for the power constraint networks. Below we describe the simple custom time synchronization design for the battery-less network over BLE.

3 Implementation

3.1 Design Requirement

Nodes in the network are in three possible states regarding to the radio activity: transmit, receive and sleep.

The Fig. 1 shows the typical communication sequence of two unsynchronized devices. If transmit of one device overlaps with receive of another, packet exchange occurs.

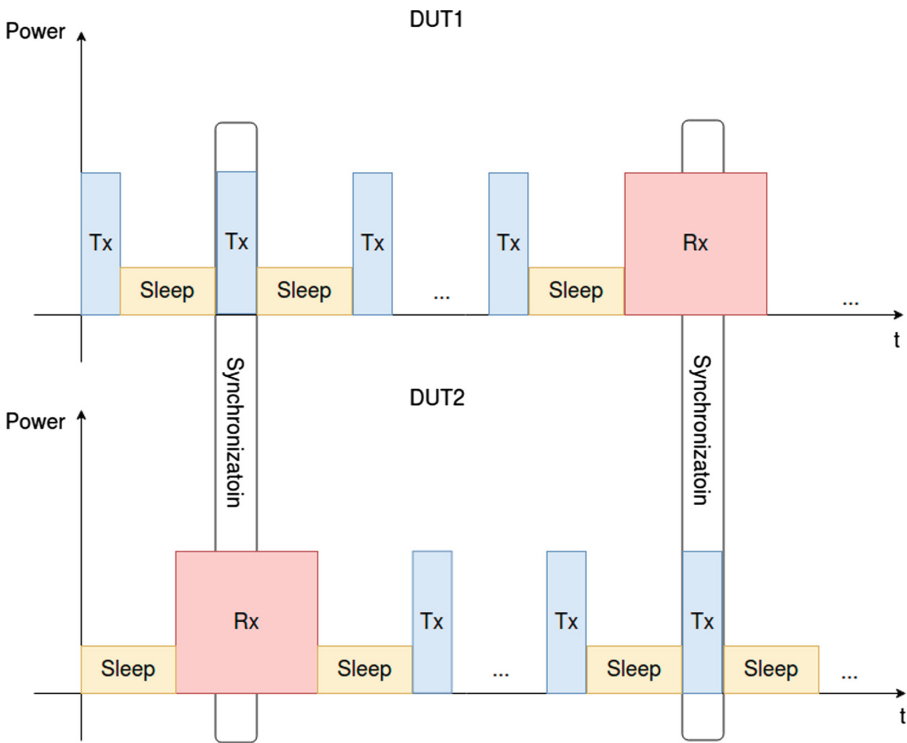


Fig. 1. Devices state changes during synchronization

Transmit and receive states consume relatively big amount of energy. Sleep is the lowest power mode in the system during which all the components are in a low-power mode. Only the sub-system responsible for the system wake-up remains active. Our aim is to get minimal power consumption that will allow to build devices without active power supply using only the ambient energy sources like the indoor lighting. To achieve this goal, we need to minimize the time of radio activity and maximize the time that devices spend in sleep. The radio activity time should be the smallest that still allows to reach reasonable communication abilities and synchronization precision. The methods that allow good synchronization precision in the distributed WSN are commonly power hungry and the trade-off exists between the power efficiency and the synchronization precision. In our case, the primary goal is to achieve the lowest possible power consumption.

3.2 Power Efficient Packet Exchange

To implement the time synchronization in a specific mesh network and make it power-efficient, we need to focus on several objectives:

- Measure existing power consumption on each of the system states, and minimize power loss in the system
- Analyze and select optimal time balance of each stage basing on the available power budget
- Implement synchronization and measure its precision and power consumption

We focused on the BLE mesh system without the actual connection events (the so-called flood mesh) that allows to populate system states across the network. In this case, we have a couple of radio activities:

1. Transmission of advertisement packet. This packet contains information about the device: its ID, name, connection ability and specific custom manufacture information. The transmitting time is relatively short: around 2 ms. Interval between the advertising packets has big influence on the average current consumption. More details are described in Sect. 4.1.
2. Scan for the advertising neighbors. This process typically takes much longer than advertisement to make sure that synchronization intersection between the nodes will happen (see Fig. 1). Scan interval is showed as solid to demonstrate the idea, but practically it should be split into shorter intervals and occur between advertising with offsets. Thus, it is effective the same but does not require to collect a lot of energy for the uninterrupted scan. More details are described in Sect. 4.2.
3. Sleep is the state with minimal power consumption. More details are described in Sect. 4.3.

3.3 Network Clock Synchronization

For implementation of a network clock we used application timer that runs from the external 32.768 XTAL clock source (20ppm). There are multiple preconditions that will cause synchronization error for this system:

1. Base tolerance 20ppm will cause 20us error every 30.6 s
2. Load capacitance mismatch of the crystal (due tolerance) and temperature variations will increase this number
3. Overflow of 24-bit counter will occur every 8.53 min and need to be handled

The actual timer ticks were used to track the uptime on each device (node). Let's define network clock as some counter that runs at the same speed as uptime on each node. Thus, knowing the difference between local uptime and network clock, nodes know network clock. At the node power-up this difference is unknown and should be adjusted during the synchronization.

Before sending an advertising packet, node writes network clock into it. The receiving device compares its own network clock with the one received from the network. The older value dominates and is used to calculate the new difference. Eventually each node knows the network clock, which is de-facto the uptime of the oldest node in the networks.

4 Measurement Results

For measurements and prototyping, the nRF52840 based board [9] was used. The measured system efficiency was defined by wasted energy to perform synchronization and communication tasks and achieved synchronization clock accuracy. The current consumption of the prototypes was measured with the specific kit (X-NUCLEO-LPM1A) that provides power to the device under test and allows to make dynamic measurements up to 50 mA with 100 kHz bandwidth. It was used to track average current consumption and energy for each zone of interest during BLE radio activities like advertisement and scanning. The baseline current consumption of the device under test (DUT) was minimized by turning off all sensors power using the power switch and turning off sensors data acquisition code in the firmware.

4.1 Advertising

Advertisement process is one of the basic BLE functions, allowing the device to identify itself, its communications ability, and transfer manufacturer-specific or custom information to the neighbors. The information in the advertisement packet is transmitted periodically. There are three dedicated channels for that. The number of channels that are used, packets length, and advertisement period define the energy that will be consumed by device during the advertising. Power consumption during the advertising is shown on the Fig. 2.

The device is configured as not-connectable, which allows to avoid radio switching to receive mode after advertising on one channel and directly switch to the next one. The measured energy for advertising on 3 channels is

$$E_{adv,3ch} = 36 \mu\text{J}. \quad (1)$$

That allows most of the BLE beacons to work for years from a single coin-cell battery. The first spike on the Fig. 2 (at around 140 ms) is caused by the wakeup

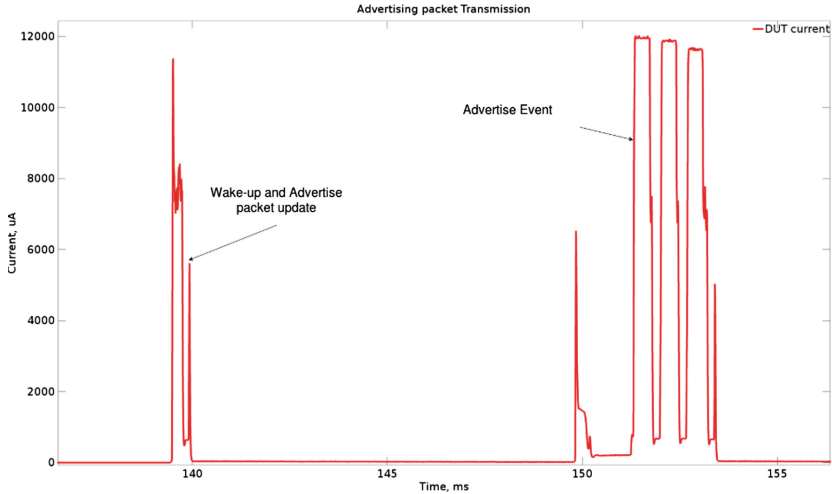


Fig. 2. Advertisement current consumption

from the application timer to dynamic update of the advertisement packet with the known network clock (system uptime plus difference). By advertisement on a single channel, it's possible to reduce the amount of the consumed energy to $16.8 \mu\text{J}$ but it decreases probability of packet delivery. The duration of transmission process on 3 channels is about 2 ms.

4.2 Scanning

After some advertisement period, where each device floods network clock to the neighbors, there is a scanning period used for listening to the other nodes and receiving their data. The scan window and interval identify the period between the sequential scans and their duration.

Minimum possible Bluetooth scan window is 2.5 ms, power consumption during that minimal scan window is shown on the Fig.3. Measured energy loss during this scan is

$$E_{scan,min} = 13.75 \mu\text{J}. \quad (2)$$

For practical usage, we could select longer scan window. Measurement of continuous scan gives us power that is needed for scan

$$P_{scan} = 22 \text{ mW}. \quad (3)$$

4.3 Sleep

To minimize board power consumption, the sensors were turned off using the power switches. Main controller (nRF52840) was put in a sleep mode. Measured power that is needed to supply device in sleep mode is

$$P_{sleep} = 6.7 \mu\text{W}. \quad (4)$$

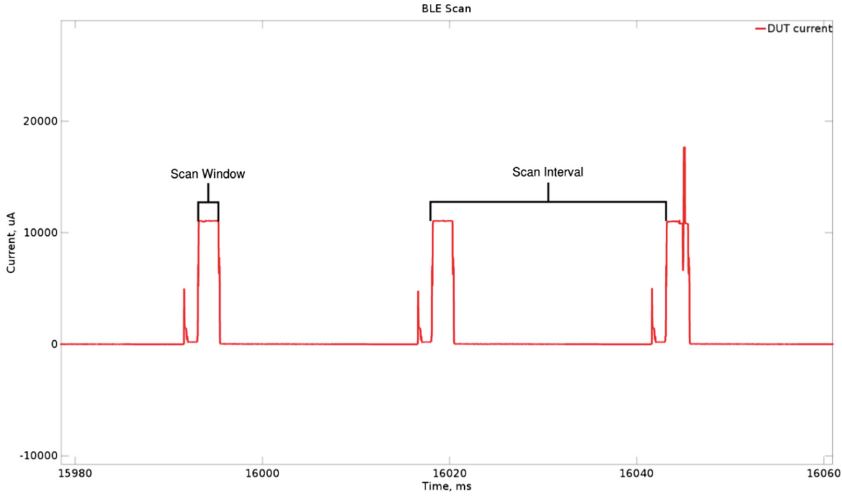


Fig. 3. Scanning current consumption (minimal scanning interval)

5 Parameters Selection

For calculations of communication parameters that will best fit our power budget we consider the following input data:

- Supply power P that is typically panel power excluding average sleep power and probably excluding power needed for the other tasks.
- The scan interval should be equal or bigger than the advertising interval, to avoid missing of the synchronization event. From power perspective we select the advertise interval equal to the scan interval in order to save energy.

Thus, we have two free parameters that should be selected: one full cycle time T and advertise/scan interval t . Its visualization is shown on Fig. 4.

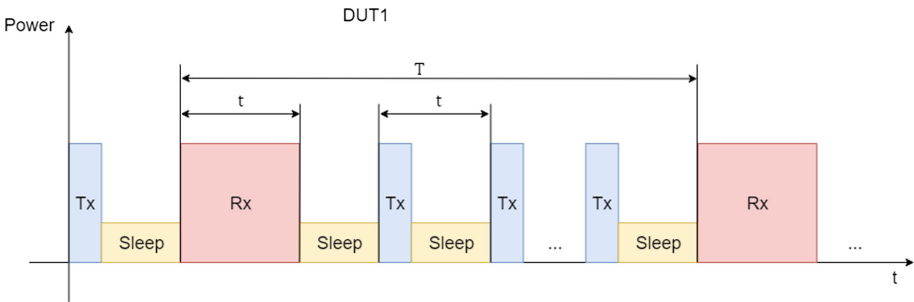


Fig. 4. Parameters visualization.

Based on these inputs, energy, consumed by device during one full cycle T is

$$E_{cons} = NE_{adv} + P_{scan}t, \quad (5)$$

where N is number of advertising packets within interval T :

$$N = \frac{T}{t} - 1. \quad (6)$$

Supply energy for one full cycle T

$$E_{supply} = PT. \quad (7)$$

To make the system function, we need to assume that the supply provides more energy than the system consumes:

$$E_{supply} > E_{cons}. \quad (8)$$

Increasing full cycle time T , we could increase excess of E_{supply} over E_{cons} as the most power consumption part (scanning) remains the same. Thus, we could find the minimal full cycle time in case $E_{supply} = E_{cons}$:

$$T_{min} = \frac{P_{scan}t^2 - E_{adv}t}{Pt - E_{adv}}. \quad (9)$$

Constraints, to have enough energy for advertising the next condition must be satisfied:

$$Pt > E_{adv}. \quad (10)$$

Here, advertising interval t is not set and could be selected depending on the actual use case and hardware restrictions. In case it could be selected, let's find the value that gives minimal T_{min} (9):

$$t_{min} = \frac{E_{adv}}{P} + \frac{E_{adv}}{P} \sqrt{1 - \frac{P}{P_{scan}}}. \quad (11)$$

First component is obvious — time required for collecting energy for advertising packet, second component is time for collection energy for scan.

Finally, minimal full cycle duration is

$$T_{min} = \frac{E_{adv} \left(2 \left[\sqrt{P_{scan}(P_{scan} - P)} + P_{scan} \right] - P \right)}{P^2}. \quad (12)$$

We assume here, that $P_{scan} > P$. Otherwise, we could scan continuously and given design is not applicable.

5.1 Example Calculation

For the example calculation we selected the solar panel EXL10-4V170 from Lightricity [6]. Lightricity’s photovoltaic energy harvesting technology has been optimised for use in low light level indoor environments and can deliver over $20 \mu\text{W}/\text{cm}^2$ at 200 lx, while still operating down to 10 lx. Combined with power management and storage in a similar way to [7], a 10 cm^2 Lightricity PV panel can provide sufficient power to all the BLE devices tested as part of this work.

Let’s assume the solar panel provides $P_{panel} = 160 \mu\text{W}$ (Lightricity panel was evaluated in [8, 10]). Excluding P_{sleep} and making some backup for variation in lighting level and panel degradation, let’s assume effective power supply energy is

$$P = 140 \mu\text{W}.$$

The advertise energy E_{adv} and scan power P_{scan} was measured in Sects. 4.1 and 4.2 correspondingly, so we can calculate the minimal advertisement/scanning interval and full clock sync duration using Eqs. (11) and (12):

$$\begin{aligned} t_{min} &= 0.57 \text{ s}, \\ T_{min} &= 179 \text{ s}. \end{aligned}$$

Number of advertising event (6) in full cycle is

$$N = 313.$$

These parameters were tested in our lab. It looks quite pessimistic for the real-time applications, but let’s highlight, that it is the initial (power-up) clock synchronization that allows nodes to schedule more efficient packet synchronization on the later stages.

6 Results and Future Steps

The synchronization precision was measured on two devices by using the logic analyzer Analog Discovery 2. The firmware uses application timer to trigger GPIO updown to light-up LED. The synchronization process can be easily visualized and measured as it is shown on Fig. 5.

Measured synchronization error is in range 2.5–20 ms. Its actual value depends on initial delay between the updates of the advertisement packet on the transmitter side; receive and process moment on the reception side. The crystal tolerance introduces additional error that increases in time during the work and cleared on the next synchronization event. Note, that the aim of this work is not to minimize the synchronization error but to make the synchronization power effective.

Analysis of the dependency between the synchronization error and the energy consumption is an open question for the future investigation. Described method is not designed to minimize the synchronization error and should be used for the initial synchronization. Achieved error (up to 20 ms) is significantly lower than the selected scan window (0.57 s). Thus, after the initial synchronization, nodes could use shorter scan window and schedule communication beforehand.

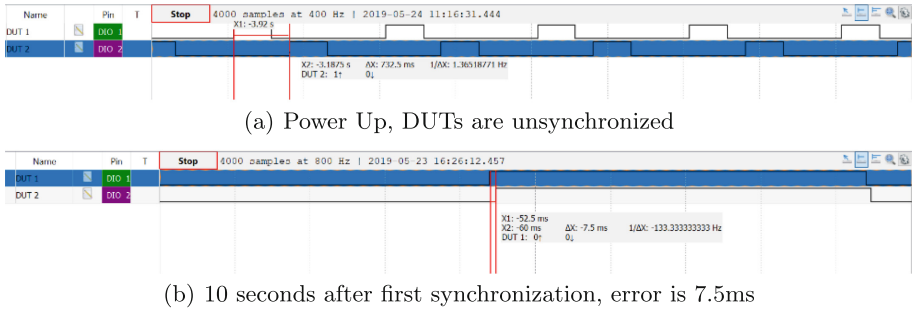


Fig. 5. Synchronization error.

7 Conclusions

Synchronized mesh network can be used in a very wide field of modern IoT application that will benefit from its ability to work with the limited power supply. In our work we showed how to implement battery-less mesh network that is using indoor light as a main power source. Synchronization of all the devices in the network allows to minimize the power loss in the system, to perform simultaneous actions in a group of the devices and register some specific events.

Provided recommendations were tested in a lab using Nordic nRF52850-based devices and efficient solar panel. Using only $140 \mu\text{W}$ from this panel, the designed clock synchronisation allows 2 nodes to be synchronized within 180 s time interval with error up to 20 ms.

The further work is expected to be done to minimize the synchronization error by implementing one of the common synchronization schemes. It is also necessary to consider the packet loss and the probability of synchronization with different parameters considering model best parameter selection by the alternative methods and this can be a subject of the further investigation.








References

1. Asgarian, F., Najafi, K.: Time synchronization in a network of bluetooth low energy beacons. In: SIGCOMM Posters and Demos, pp. 119–120 (2017). <https://doi.org/10.1145/3123878.3132007>
2. Roche, M.: Time Synchronization in Wireless Networks (2015). https://www.cse.wustl.edu/~jain/cse574-06/ftp/time_sync/index.html
3. Hua, Y., Xin, Y., Kangli, L., Shenghui, W.: Research overview of clock synchronization in wireless sensor network. J. Comput. Sci. Appl. Inform. Technol. **3**(1), 1–10 (2018). <https://doi.org/10.15226/2474-9257/3/1/00128>
4. TinyOS Documentation Wiki main page. http://tinyos.stanford.edu/tinyos-wiki/index.php/Main_Page
5. Rahamatkar, S., Agarwal, A.: A reference based, tree structured time synchronization approach and its analysis in WSN. Int. J. Ad hoc, Sens. Ubiquit. Comput. **2**(1) (2011). <https://doi.org/10.5121/ijasuc.2011.2103>

6. Lightricity ExcellLight EXL10-4V170. <https://lightricity.co.uk/excellight-exl10-4v170>
7. Yue, X., et al.: Development of an indoor photovoltaic energy harvesting module for autonomous sensors in building air quality applications. *IEEE Internet of Things J.* **4**(6), 2092–2103 (2017). <https://doi.org/10.1109/JIOT.2017.2754981>
8. Lightricity solar panels evaluation. <https://github.com/softserveinc-rnd/blog/tree/master/solar-panels>
9. BLE5 node for prototyping. <https://github.com/softserveinc-rnd/ble5-node-prototype>
10. Testing indoor light as a power source. <https://www.softserveinc.com/en-us/blog/testing-in-door-light-as-a-power-source/>
11. Rheinlander, C.C., Wehn, N.: Precise synchronization time stamp generation for Bluetooth low energy. *IEEE Sens.* (2016). <https://doi.org/10.1109/icsens.2016.7808812>
12. Sridhar, S., Misra, P., Gill, G.S., Warrior, J.: CheepSync: a time synchronization service for resource constrained bluetooth le advertisers. *IEEE Commun. Mag.* **54**(1), 136–143 (2016). <https://doi.org/10.1109/mcom.2016.7378439>
13. Dian, F.J., Yousefi, A., Somaratne, K.: A study in accuracy of time synchronization of BLE devices using connection-based event. In: 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (2017). <https://doi.org/10.1109/iemcon.2017.8117156>
14. Nayyer, A., Nayyer, M., Awasthi, L.K.: A comparative study of time synchronization protocols in wireless sensor network. *Int. J. Comput. Appl.* **36**(11), 13–19 (2011)



Classifying Teachers' Self-reported Productivity, Stress and Indoor Environmental Quality Using Environmental Sensors

Johanna Kallio^(✉) , Elena Vildjiounaite , Vesa Kyllönen,
Jussi Ronkainen , Jani Koivusaari , Salla Muuraiskangas ,
Pauli Räsänen, Heidi Similä , and Kaisa Vehmas 

VTT Technical Research Centre of Finland, Kaitoväylä 1, 90571 Oulu, Finland
johanna.kallio@vtt.fi

Abstract. Considering that urban people spend majority of their time indoors, buildings should support health and productivity. However, quite commonly unsatisfactory Indoor Environmental Quality (IEQ) causes environmental stress, which can lead to adverse health outcomes and reduced productivity. First step to enable automatic environmental control is to recognise environmental conditions that can negatively influence each individual. To this end, we developed (1) multi-sensor IEQ monitoring system to measure objectively environmental quality; (2) mobile application to collect subjective evaluation of productivity, stress and IEQ data; (3) machine learning method to use IEQ data to distinguish between positive and negative self-reports of test subjects. Experimental results with real life data, collected in four classrooms of Finnish elementary school during 18 weeks, show that IEQ sensor data allows to classify with fairly high accuracy perceptions of teachers regarding their work productivity (91%), stress (81%) and IEQ (92%). This result was achieved in person-specific training (*i.e.*, model of each individual was trained using only his/her data), whereas accuracy of leave-one-person-out approach was notably lower. These results suggest that perception is personal and some individuals are more sensitive to environmental stressors.

Keywords: Indoor environmental quality · Machine learning · Classification · Productivity · Stress · Self-report

1 Introduction

Recent advancements in sensor and mobile technology have enabled real time and long-term data collection from our surroundings and human behavior without causing disruptions on our daily routines. Specifically, environmental sensors and smartphones, which can capture multimodal data, have the potential to identify environmental risk factors and provide information to improve indoor environmental quality (IEQ) and wellbeing. IEQ relates to the interaction between physical, chemical and biological factors indoors, such as air quality pollutants, temperature, relative humidity, lighting and sound level, and their influence on quality of life [1, 2]. These factors, however, can cause stress and can be considered as environmental stressors, in other words,

factors that appear in our environment and that can trigger an individual's stress response [3].

During the past few decades, insufficient IEQ has been shown to cause adverse health effects and to reduce cognitive performance and productivity [4, 5]. Lamb et al. [6] conducted an 8-month field study in real office environment and found that perceived thermal discomfort, noise annoyance and lighting discomfort were associated with reduced work performance and objectively measured cognitive performance. In most cases, environmental stress reduced self-reported work performance and measured cognitive performance by 2.4 to 5.8%. However, this 8-month study did not include objective measurements of the IEQ factors; instead, the environmental stressors and the work performance were measured with a subjective self-reporting questionnaire.

In another study [7] that evaluated environmental parameters, IEQ perception and productivity showed that the optimal productivity was achieved when subjective evaluation of thermal condition was ranked as "neutral" or "slightly cool". The experiment was conducted in a controlled office environment and the measured IEQ factors included temperature, relative humidity, CO₂, lighting and noise level. A recent review study [8] also consolidated that both thermal and indoor air quality discomfort influences the work performance. The reviewed experiments lasted from 2 h to 36 days and they were carried out in controlled environments or climate chambers, where IEQ factors can be changed one by one and more accurately than in real life. For example, in many studies researchers changed only temperature and observed how thermal changes affect either perceived productivity or ability of test subjects to perform certain tasks (*e.g.*, measured changes in typing speed or number of errors). In other studies researchers increased/decreased ventilation rate and measured productivity, but they did not measure IEQ using sensors.

Moreover, providing the environmental comfort is complicated by the fact that people differ greatly in their sensitivity and stress-proneness. Some individuals are more sensitive to the nature and severity of (environmental) stressor experiences that may strain wellbeing and performance [3, 9]. For instance, some individuals are more prone to sound and temperature changes [10], whereas females are more likely to express lower satisfaction with IEQ factors than males [11]. Several previous studies have found that person-specific stress classifiers have achieved better accuracy than general stress classifiers trained on data of many users [12].

Measuring IEQ in real building and real life conditions should be non-invasive and continuous. Advancements in wireless technologies and electronics have enabled the development of low-cost, low power and multifunctional environmental sensors, which do not require installation or maintenance efforts from end users. These unobtrusive sensors can provide new means to collect continuous and real-time data about indoor environments, while machine learning methods allow to recognise, which objective environmental parameters, obtained from these sensors, ensure personal comfort vs. cause discomfort. In this initial study into recognition of personal perceptions of productivity, stress and IEQ on the basis of IEQ sensor data, we used supervised learning.

In supervised learning, data with known labels are used to train the classifier to distinguish between classes. Especially Support Vector Machine (SVM) classifier has

been successfully used in various domains, including stress recognition [12]. SVMs use nonlinear mapping to represent data in a high dimension that is typically much higher than the original feature space [13]. Then a linear classifier can be applied to separate the classes. The hyperplane separating the classes is selected so that there is maximum distance between the training data classes.

The main contribution of this paper is twofold; first, we introduce a continuous IEQ monitoring system to collect and process indoor environmental measures using commercial sensors. Second, we present experimental results of using objective IEQ data to classify personal perceptions of teachers regarding “negative” or “positive” productivity, stress and IEQ. In experiments with real life data of four teachers, collected during 18 weeks, the proposed system achieved the following average accuracies: productivity - 91%, stress - 81% and IEQ - 92%. To the best of our knowledge, our study is the first one, which has collected long-term real life IEQ data with multiple environmental sensors in school environment and presented a machine learning approach to use these data to classify personal perceptions of IEQ, work productivity and stress.

2 Methods

2.1 Participants

Our study recruited participants from Finnish elementary schools. We decided to choose a school as a pilot facility since EFA [14] has found that 64 million students and 4.5 million teachers are affected by insufficient IEQ in Europe. We distributed our study advertisement in the form of press release. Potential participant schools filled out a questionnaire to determine eligibility. Our inclusion criteria for the schools were (a) elementary school where students were spending most of the day in one classroom, (b) perceived indoor air quality problems, (c) no major renovations in sight, and (d) close proximity to offices of VTT Technical Research Centre of Finland Ltd. (responsible for organizing the study). Inclusion criteria for the participating teachers were (a) Android phone in use, (b) commits to wear wrist-devices for data collection during the school day, and (c) commits to answer self-reports daily. We targeted only Android phone users because the application used for collecting self-report data was developed for Android OS. The study participants wore a wrist device to monitor continuous activity and heart rate data during working hours. However, the physiological data was not used in the analyses reported in this paper.

We decided on the pilot school based on the inclusion criteria. The researchers went to present the study to school for recruiting voluntary teachers. The volunteers ($N = 4$) attended information session, where they were able to present questions related to the study and informed consent that was sent beforehand. After the information session, one volunteer changed school and another volunteer was recruited and informed separately. The average age of the voluntary teachers was 43.5 years (standard deviation 15.6) and all of them were females. The signatures were collected to the consents and relevant Android applications were installed during the information session. The study protocol was reviewed and accepted by the Ethics Review Board of VTT Technical Research Centre of Finland Ltd.

2.2 Continuous Environmental Monitoring

The data was collected during an 18-week pilot via IEQ monitoring systems installed in four pilot classrooms and an Android self-reporting application used by the teachers. The high-level data collection set-up is presented in Fig. 1. All pilot classrooms were equipped with a set of commercial MFC-LW12CO2 sensor devices measuring temperature (T), relative humidity (H), air pressure (P), carbon dioxide (CO₂), indoor air quality index (IAQ index) and ambient luminosity (LUX) [15]. IAQ index represents breath-VOC (b-VOC) concentration for the most important compounds in exhaled breath of humans and is output of a proprietary algorithm in the VOC gas sensor by the sensor vendor. It should be noted that we did not verify the accuracy specifications of the measurement devices provided by manufacturers. All the sensor devices were installed at about 1.1 meters above the floor level next to teacher’s post in the classroom according to national legislative recommendations [16]. The sampling rate was once per 15 min.

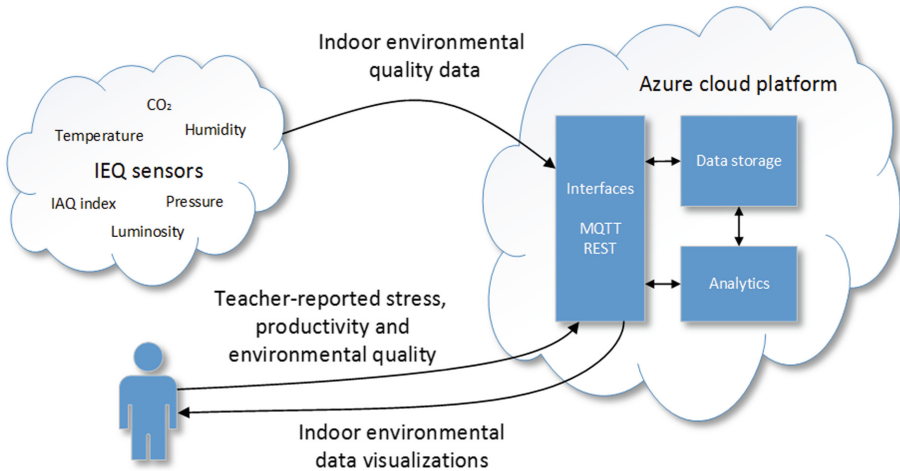


Fig. 1. Data collection set-up in all four pilot classrooms.

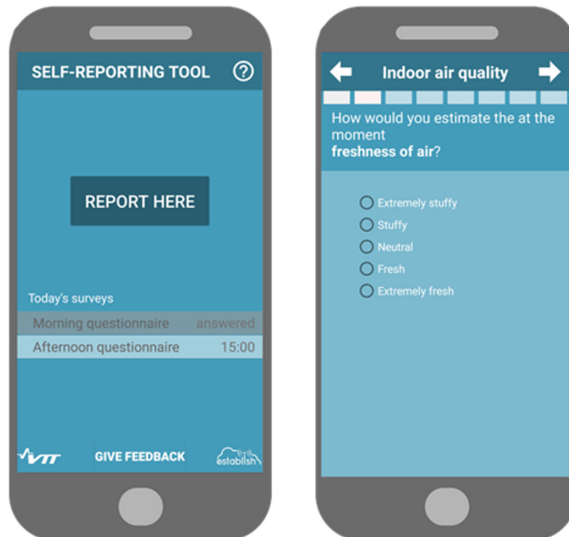
The IAQ sensor devices at the pilot site utilized LoRaWAN for communication. A commercial LoRaWAN gateway deployed at the site aggregated the sensor device data and communicated to a LoRaWAN server running in the cloud. The LoRaWAN server in turn relayed the sensor data intracloud via MQTT to Azure IoT Hub. All collected sensor data were stored in Azure Table storage. Furthermore, teachers were given a possibility to view measurement data via visualizations. Table 1 sums up the sensor device and app capabilities that were installed per classroom and utilized in the pilot respectively.

Table 1. Details on the collected data

Data source	Parameters	Measurement range	Sampling	Data connection
MCF sensor device	Temperature (T)	-10...60 °C	Every 15 min	LoRaWAN + MQTT/IP
	Relative humidity (H)	0...100%		
	Air pressure (P)	300...1100 hPa		
	CO ₂	300...5000 ppm		
	Ambient lighting (LUX)	0.01...80000 lx		
	IAQ index	0...500 (b-VOC sensor 500...50000 ppb)		
Android self-reporting app	Self-reported productivity, stress and IEQ	1...5	Twice on weekdays	REST/HTTPS/IP

2.3 Self-reporting Application

Subjective evaluations of productivity, stress, IEQ and any related symptoms were collected from participating teachers via a self-reporting application on the teachers' Android phones. Figure 2 shows screenshots of the application.

**Fig. 2.** Screenshots of the self-reporting application

The user logged in once at the beginning of the measurement period. From then on, the app reminded her to answer the questionnaire at given times of day on weekdays, and provided a button for spontaneous reporting of symptoms or environmental quality. The questionnaire reminders were implemented as Android notifications. There were two time-triggered questionnaires on each school day (morning and afternoon questionnaires). Timing of the questionnaires was selected based on the schedules of the teachers. Morning questionnaire popped out at 7:30 and was active until 11:30. Afternoon questionnaire popped out at (noon) 12:00 and was active until 17.00 on Mondays, Tuesdays and Thursdays. On Wednesdays and Fridays afternoon questionnaire popped out at 10:30 and was active until 16.30. This schedule allowed the teachers to choose a convenient moment to answer, but reduced the probability to collect outdated answers.

The morning questionnaire asked about how rested the teachers felt and perceived environmental quality. Perceived IEQ was asked with one general question and five additional items about air freshness, odor intensity, temperature, humidity, and noise level. The afternoon questionnaire asked perceptions of productivity, stress, IEQ and related symptoms, actions during the day for improving environmental quality, pupil restfulness, and pupil concentration. The answer scale was 5-point with answer options (1) Strong negative/very bad, (2) Negative/bad, (3) Neutral/acceptable, (4) Positive/good, (5) Strong positive/Very good. Analyses in this paper focus on perceived productivity, stress and general IEQ. In addition, the application allowed the teachers to report any perceived issue (*e.g.*, IEQ problem, perceived productivity etc.) any time they wanted, also on 5-point scale.

The users' answers were stored in MongoDB as one document per report or questionnaire item (*e.g.*, one question). The answers pertaining to one questionnaire were determined by their shared timestamp. Storing the individual items rather than the whole questionnaire as a single document added flexibility to data storage and fetching the answers to specific questions only, regardless of whether they were answered as a part of a questionnaire or as a spontaneous user-triggered report.

The client app got the assets from the server and rendered them into user interface elements. Each user's answers were tagged with a pseudo-anonymization code that the application received during first login. The user then answered the questionnaires within their validity windows, as reminded. If the user missed the window, the questionnaire was no longer available for that day. At the end of the questionnaire or report, the user provided their location at the time of answering, *e.g.*, "Room A123". The room codes were defined by the user, and pseudo-anonymized using DES encryption before sending to the server. The application server also provided a REST API for updating the assets, monitoring the answering activity, and fetching the recorded answers.

2.4 Data Analysis

The goal of data analysis was to distinguish between "bad" and "very bad" personal perceptions and more positive perceptions. We performed 2-class classification, *i.e.*, for stress "bad" class included "negative" and "very negative" self-reports, and "OK" class included "neutral", "positive" and "very positive" self-reports. Similarly, for productivity classification "bad" class included "much less productively than usually" and

“less productively than usually” answers, while “OK” class included all other answers. For IEQ perception “bad” class included “bad” and “very bad” answers, and “OK” class included all other answers.

We employed fully supervised classification in this study: we used SVM classifier and its implementation in Scikit-learn Python library. We trained SVM to map sensor data into self-reports of each type, in other words to predict whether answer for each question type (productivity, stress, IEQ) will be “bad” or “OK”. In the experiments, we compared two different approaches: general and person-specific learning. In general learning, data of the target person (*i.e.*, data of the current test subject) is not used in any way in training; in person-specific learning, on the contrary, only data of the target person is used in training.

In this study, we used leave-one-person-out protocol for assessing accuracy of general approach. In leave-one-person-out protocol, test dataset contains all data of one test subject, and training dataset contains data of all other test subjects. This process is repeated for all subjects. For person-specific learning, we used leave-one-self-report-out protocol: test dataset contains one self-report and the corresponding sensor data; training dataset contains all other data of this person. This process is repeated for all self-reports of this person, and then the same procedure is repeated for all other persons.

General and person-specific approaches differ only in a way how training and test datasets are created; all other steps are similar. First, we extracted data features from training data. For feature extraction and selection of the most relevant features we used tsfresh Python library. It automatically extracts over 300 features, such as minimum, maximum, energy, quantiles, various data change metrics, and uses self-reports from the training data to assess false discovery rate of each feature and to select most relevant features this way (for more details see [17]). In the experiments, presented in this paper, we extracted features from time-series data within two-hour-long time window preceding each self-report (all data samples and self-reports are accompanied with timestamps in our system). Then we trained SVM using extracted features and corresponding labels.

For testing, we extract the same features from the test data, classify each test sample with SVM and compare SVM output with the corresponding self-report. We estimate test accuracy according to the following measures:

$$Total\ accuracy = \frac{N_{correctBad} + N_{correctOK}}{N_{bad} + N_{OK}} \quad (1)$$

$$True\ bad = \frac{N_{correctBad}}{N_{bad}} \quad (2)$$

$$True\ OK = \frac{N_{correctOK}}{N_{OK}} \quad (3)$$

where $N_{correctBad}$ is the number of correctly classified “bad” answers; $N_{correctOK}$ is the number of correctly classified “OK” answers, N_{bad} N_{OK} are numbers of “bad” and “OK” answers, respectively.

3 Results

3.1 Self-report Dataset

Collected self-report dataset contained 569 answers with average of 8 weekly answers per person (minimum 3, maximum 12 weekly reports per person). “Bad” stress class was reported on average in 35% of days, “bad” productivity class was reported on average in 18% of days, “bad” environmental quality was reported on average in 31% of days. Test subjects notably differed in their perceptions of productivity, stress and IEQ. The most satisfied subject reported “bad” productivity on 4% of days, “bad” stress on 17% of days and “bad” environmental quality on 5% of days. The most dissatisfied subject reported “bad” productivity on 48% of days, “bad” stress in 70% of days and “bad” environmental quality in 70% of cases.

“Bad” self-reports of different types co-occurred (*i.e.*, both bad problems) fairly frequently. On average, “bad” stress was reported on the same day in 88% of cases when “bad” productivity was reported. Negative stress and low productivity correlated with poor IEQ: on average, “bad” IEQ and “bad” stress were reported on the same day in 54% of days when “bad” productivity was reported. “Bad” IEQ was reported on the same day in 89% of days when “bad” stress was reported and 90% of the days when “bad” productivity was reported.

3.2 Productivity Classification

Table 2 presents classifications accuracies for two classes of productivity, achieved with different sensor combinations. Here, THP stands for “temperature, relative humidity, air pressure” sensor combination.

Table 2. Productivity classification accuracies of person-specific models with different sensor combinations

Sensors	Total accuracy	True bad	True OK
THP, CO ₂ , IAQ index	0.89	0.75	0.92
THP, CO ₂	0.89	0.75	0.92
THP, IAQ index	0.91	0.71	0.96
THP	0.90	0.74	0.95
THP, CO ₂ , IAQ index, LUX	0.89	0.68	0.94
THP, LUX	0.80	0.41	0.88
THP, CO ₂ , LUX	0.89	0.68	0.94
THP, IAQ index, LUX	0.80	0.41	0.88

The highest productivity classification accuracy (total accuracy of 91%) was achieved with THP and IAQ index data. For comparison, result of the general approach with sensor combination “THP, CO₂, IAQ index, LUX”: Total accuracy = 0.66; True bad = 0.07; True OK = 0.84.

3.3 Stress Classification

Table 3 presents classifications accuracies for two classes of stress, achieved with different sensor combinations.

Table 3. Stress classification accuracies of person-specific models with different sensor combinations

Sensors	Total accuracy	True bad	True OK
THP, CO ₂ , IAQ index	0.80	0.68	0.85
THP, CO ₂	0.80	0.59	0.91
THP, IAQ index	0.75	0.56	0.86
THP	0.75	0.56	0.86
THP, CO ₂ , IAQ index, LUX	0.79	0.70	0.84
THP, LUX	0.79	0.62	0.88
THP, CO₂, LUX	0.81	0.70	0.87
THP, IAQ index, LUX	0.79	0.62	0.88

The accuracy for stress classification was highest (total accuracy of 81%) with THP, CO₂ and LUX data. However, differences in classification accuracies with various sensor data combinations were rather small. For example, the total classification accuracy based on THP, CO₂ and IAQ index data was the same (80%) than the total classification accuracy based on THP and CO₂ data. For comparison, result of the general approach with sensor combination “THP, CO₂, IAQ index, LUX”: Total accuracy = 0.37; True bad = 0.18; True OK = 0.47.

3.4 Indoor Environmental Quality Classification

Table 4 presents classifications accuracies for two classes of IEQ, achieved with different sensor combinations.

Table 4. IEQ classification accuracies of person-specific models with different sensor combinations

Sensors	Total accuracy	True bad	True OK
THP, CO ₂ , IAQ index	0.90	0.78	0.94
THP, CO₂	0.92	0.80	0.95
THP, IAQ index	0.90	0.72	0.95
THP	0.91	0.77	0.95
THP, CO ₂ , IAQ index, LUX	0.90	0.77	0.94
THP, LUX	0.90	0.75	0.95
THP, CO ₂ , LUX	0.90	0.77	0.94
THP, IAQ index, LUX	0.89	0.73	0.93

The accuracy IEQ classification (total accuracy of 91%) was highest when using THP and CO₂ data. However, differences in IEQ classification accuracies with various sensor combinations were small. For comparison, results of the general approach with sensor combination “THP, CO₂, IAQ index, LUX”: Total accuracy = 0.18; True bad = 0.12; True OK = 0.19. These results show that perception of environmental quality also notably depends on an individual.

4 Discussion

4.1 Principal Findings

In this paper, we presented a multi-sensor IEQ monitoring system to measure objective indoor environmental data and a mobile application to collect perceived productivity, stress and IEQ data. Our aim was to study how accurately the IEQ measures could identify conditions of self-reported productivity, stress and IEQ, and which of the measured IEQ data were most accurate in identifying those conditions. For classifying “bad” or “OK” stress, productivity and IEQ, we used SVM classifier, which has been successfully used for classification in earlier studies [12]. Our data analysis indicated that relatively high accuracies could be achieved using person-specific SVM models to classify self-reported productivity, stress and IEQ. For comparison, results of general SVM models for two-class classification of productivity, stress and IEQ were worst in all cases.

Depending on the IEQ sensor data combinations, the total person-specific classification accuracies for perceived productivity varied from 81% to 91%. The best IEQ sensor data combination for work productivity classification seemed to be temperature, relative humidity, air pressure and IAQ index data. This finding is consistent with previous studies showing that both thermal and indoor air quality discomfort have negative effects on work performance [8]. On the other hand, the differences in productivity classification accuracies (total accuracies from 89% to 91%) using temperature, relative humidity, air pressure, CO₂ and IAQ index data were insignificant. However, ambient luminosity measures did not have significant effect on the classification results, even though lighting discomfort has found to be associated with objectively measured performance in earlier studies [6]. Probably, this is due to the proper lighting conditions in our study, and thereby the teachers did not experience lighting discomfort.

The total accuracies of stress classification varied from 75% to 81% for the applied combinations of measurements. The best stress classification accuracy was achieved with combination of temperature, relative humidity, air pressure, CO₂ and ambient luminosity sensor data. On the contrary, the lowest total accuracy (75%) was achieved with temperature, relative humidity, air pressure and IAQ index data combination, which again was the best data combination for the productivity classification (91%). The total classification accuracies for perceived IEQ varied from 89% to 92%. For IEQ, the highest classification accuracy (92%) was achieved using THP and CO₂ sensor data, but almost the same accuracy was achieved without CO₂ measures (91%). Overall, most of the presented sensor combinations result in a satisfactory accuracy in

identifying the relation to reported stress, productivity, and IEQ. According to the selected use case, recommendations for the preferred set of sensors can be identified.

In the self-reports, feedback on simultaneous “bad” conditions co-occurred rather frequently. For instance, “bad” stress was reported on the same day in 88% of cases when “bad” productivity was reported. Whether highly stressed feeling is caused by low productivity or vice versa requires further studies, but in any case, high stress and low productivity correlated with poor IEQ. This result is in line with the findings of numerous other studies, which have indicated that bad IEQ negatively affects concentration and productivity [4, 6, 8].

4.2 Individual Differences

To better understand why person-specific learning achieved higher classification accuracies than general models, we calculated correlations between different sensor features and “bad” self-reports. These correlations revealed notable differences between features, affected perception of productivity, stress and IEQ by different persons.

First, the features with the highest correlation coefficients (over 0.4) were extracted from different sensor types. Perceived productivity of *Teacher One* was highly correlated only with pressure features; perceived productivity of *Teacher Two* was highly correlated with IAQ index and temperature features; perceived productivity of *Teacher Three* was highly correlated with temperature and CO₂ features; perceived productivity of *Teacher Four* was highly correlated with humidity, temperature and pressure features. Perceived stress of *Teacher One* was highly correlated with temperature and CO₂ features; perceived stress of *Teacher Two* was highly correlated only with temperature features; perceived stress of *Teacher Three* was highly correlated with pressure and temperature features; perceived stress of *Teacher Four* was highly correlated only with pressure features. Even perception of IEQ was person-dependent: *Teacher One* was mainly affected by humidity; *Teacher Two* was mainly affected by temperature; *Teacher Three* was affected by temperature and CO₂; for *Teacher Four* high correlation between environmental quality and sensor data was observed only with pressure features.

Second, although most often high correlations were observed between “bad” self-reports and features reflecting data changes (standard deviation, sum over the absolute value of consecutive changes, data trends etc.). In some cases high correlations were observed also between “bad” self-reports and absolute values: for *Teacher Two* perception of environmental quality highly correlated with maximum, mean and median features of temperature data, and for teacher four productivity highly correlated with longest consecutive subsequence of temperature values above temperature mean.

The higher accuracies of person-specific learning reflect the fact that people differ in their stress-proneness [3] and sensitivity to environmental stressors [10]. This finding is also congruent with a previous review study [12] showing that algorithms trained using data of a target person only; usually achieve notably higher accuracies than algorithms trained using data of other persons.

4.3 Limitations and Future Work

Due to small number of participants ($N = 4$), these findings cannot be generalized too widely. In addition, all participants were females who are more likely to show dissatisfaction with IEQ than males [11]. Thus, our results should be further studied and confirmed in populations that are more diverse and in different working environments. School environment is challenging, because we cannot for example confirm whether negative stress is caused by pupils, or environmental variables. We did not either consider psychosocial factors, which can influence the subjective evaluation of IEQ [18].

Our future work will focus on collecting a larger amount of multimodal data for longer time in various working spaces to study long-term environmental stress factors and the effects of these factors on individuals. In general, we are interested in interaction between humans, their behavioral patterns and working environment, and plan to provide personal suggestions or environmental adjustments based on different working tasks and individual needs.

We are also interested to study, how to combine our work with the work on energy efficiency of the buildings. Employees' annual salaries exceed the building operation and rental costs by a factor of up to 25 [19], so reduced productivity due to unsatisfactorily IEQ can lead to financial losses of employers and society even if energy costs decrease. Currently, overall cost of presenteeism (*i.e.*, employees who are present at work but work with reduced productivity due to various reasons) is estimated to be £15 billion per year in the UK alone [19]. One approach to this problem could be to augment behavior change solutions to promote energy efficiency (see *e.g.*, [20]) with capability to predict personal satisfaction and productivity: for example, behavior change application may suggest the user to reduce energy consumption only when this reduction is not likely to cause negative consequences.

5 Conclusions

This work introduced a multi-sensor IEQ monitoring system to measure indoor environmental factors continuously, a mobile application to collect subjective evaluation of productivity, stress and IEQ data and a machine learning approach to distinguish between positive/negative perceptions of these aspects on the basis of IEQ sensor data. We carried out an 18-week pilot with four teachers of real elementary school in Northern Finland and demonstrated the feasibility of the multi-sensor IEQ monitoring system for long-term data collection and classification of teachers' self-reports. Results indicated that objective IEQ measures can classify teachers' self-reported productivity, stress and IEQ with reasonable accuracies (over 80%) using SVM classifier. However, higher accuracies of person-specific than general approaches indicated that sensitivity to environmental factors is highly individual and should be considered when supporting health and productivity. Moreover, the findings let to extrapolate, that improving the IEQ could possibly lead to improved wellbeing in form of decreased stress, and to increased productivity. These potential advantages, combined with general health benefits for the society, provide a motivation for the future building design for enhanced indoor environments.

Acknowledgements. The authors are grateful to the pilot participants for their active and significant role in the pilot. Special thanks to Mr. Jari Rehu, Mr. Severi Olsbo and our former colleague Mr. Dan Bendas. The study was implemented in collaboration with ESTABLISH and SCOTT projects. ESTABLISH (ITEA3 15008) was supported by Business Finland and VTT Technical Research Centre of Finland, and the pilot was designed in collaboration with UniqAir, CGI, and InspectorSec. SCOTT (Secure COnnected Trustable Things) has received funding from the Electronic Component Systems for European Leadership Joint Undertaking under grant agreement No 737422. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme, and from Austria, Spain, Finland, Ireland, Sweden, Germany, Poland, Portugal, Netherlands, Belgium and Norway.


References

1. Kraus, M.: Exploring determining factors of indoor environment quality (IEQ). In: 18th International Multidisciplinary Scientific GeoConference (SGEM 2018), Sofia, vol. 18, pp. 701–706 (2018)
2. Steinemann, A., Wargocki, P., Rismanchi, B.: Ten questions concerning green buildings and indoor air quality. *Build. Environ.* **112**, 351–358 (2017)
3. Chrousos, G.P.: Stress and disorders of the stress system. *Nat. Rev. Endocrinol.* **5**(7), 374 (2009)
4. Fisk, W.J.: Health and productivity gains from better indoor environments and their relationship with building energy efficiency. *Annu. Rev. Energy Env.* **25**(1), 537–566 (2000)
5. Arif, M., Katafygiotou, M., Mazroei, A., Kaushik, A., Elsarrag, E.: Impact of indoor environmental quality on occupant well-being and comfort: a review of the literature. *Int. J. Sustain. Built Environ.* **5**(1), 1–11 (2016)
6. Lamb, S., Kwok, K.C.: A longitudinal investigation of work environment stressors on the performance and wellbeing of office workers. *Appl. Ergon.* **52**, 104–111 (2016)
7. Geng, Y., Ji, W., Lin, B., Zhu, Y.: The impact of thermal environment on occupant IEQ perception and productivity. *Build. Environ.* **121**, 158–167 (2017)
8. Wargocki, P., Ten Wyon, D.P.: questions concerning thermal and indoor air quality effects on the performance of office work and schoolwork. *Build. Environ.* **112**, 359–366 (2017)
9. Lan, L., Wargocki, P., Lian, Z.: Quantitative measurement of productivity loss due to thermal discomfort. *Energy Build.* **43**(5), 1057–1062 (2011)
10. Clausen, G., Wyon, D.P.: The combined effects of many different indoor environmental factors on acceptability and office work performance. *HVAC&R Res.* **14**(1), 103–113 (2008)
11. Kim, J., de Dear, R., Candido, C., Zhang, H., Arens, E.: Gender differences in office occupant perception of indoor environmental quality (IEQ). *Build. Environ.* **70**, 245–256 (2013)
12. Alberdi, A., Aztiria, A., Basarab, A.: Towards an automatic early stress recognition system for office environments based on multimodal measurements: a review. *J. Biomed. Inform.* **59**, 49–75 (2016)
13. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*, 2nd edn. Wiley, New York (2001)
14. EFA. Indoor air quality. <http://www.efanet.org/air-quality/indoor-air-quality>. Accessed 03 Sept 2019
15. MCF-LW12CO2. <https://www.mcf16.com/prodotto/mcf-lw12co2/>. Accessed 03 Sept 2019

16. Finnish Ministry of Social Affairs and Health: Decree of the Ministry of Social Affairs and Health on Health-related Conditions of Housing and Other Residential Buildings and Qualification Requirements for Third-party Experts (545/2015). <https://www.finlex.fi/en/laki/kaannokset/2015/en20150545.pdf>. Accessed 07 June 2019
17. tsfresh documentation. <https://buildmedia.readthedocs.org/media/pdf/tsfresh/latest/tsfresh.pdf>. Accessed 03 Sept 2019
18. Finell, E., et al.: The associations of indoor environment and psychosocial factors on the subjective evaluation of indoor air quality among lower secondary school students: a multilevel analysis. *Indoor Air* **27**(2), 329–337 (2017)
19. Al, H.Y., Arif, M., Kaushik, A., Mazroei, A., Katafygiotou, M., Elsarrag, E.: Occupant productivity and office indoor environment quality: a review of the literature. *Build. Environ.* **105**, 369–389 (2016)
20. GAIA project. <http://gaia-project.eu/index.php/en/gaia-objectives/>. Accessed 03 Sept 2019



User Requirements for the Design of Smart Homes: Dimensions and Goals

Michaela R. Reisinger¹ , Sebastian Prost², Johann Schrammel¹,
and Peter Fröhlich¹

¹ Austrian Institute of Technology, Giefinggasse 4, 1210 Vienna, Austria
michaela.reisinger@ait.ac.at

² Newcastle University, Newcastle upon Tyne NE1 7RU, UK

Abstract. The ‘Smart Home’ is a strongly technology-driven field. While user-centered requirements have been reported for specific features, a considerable gap persists for design based on an everyday home context and the social and emotional nature of the home. To address this, we identify specific leverage points and functionalities for energy-efficiency and smart control in a domestic context. A three-step investigation of user requirements, employing cultural probing, participatory design fiction and focus groups allowed us to progress from the home context “as-is” towards a blending of requirements with technological solutions. Our results highlight the home as a complex construct imbued with organically grown routines and individual needs, values and emotions. Based on empirical, real-user data we present features and system expectations that address this multifaceted overall picture. This paper advises the design process of future smart home solutions in two facets: First, we point out several *design dimensions*, namely time, space, relations, individual factors, and values that allow design for a heterogeneity of users and situations. Second, we derive specific *design goals* to highlight directions of smart home system design: design for control, low effort, integration, evolvability, identity, sociability, and benefits.

Keywords: Smart home · User requirements · Design implications

1 Introduction

“Smartness” is infectious: it is extending above and beyond specific devices, as well as contexts – including the home [1, 2]. Smart homes have become technologically available and feasible [3, 4], the interest in smart homes in research and industry is growing exponentially [1, 5], yet, their appeal to the general population is limited [2]: The “socio-economic and technological constellation” [6, p. 236] necessary for their success has not yet been met. Most visions of smart homes are not based on the context of a real home, but depicted as sterile, empty spaces at odds with home realities [1], which are not able to comprehend the emotionally laden “invisible boundaries” that form the home context. Highly individual, diverging combinations of items pose another home-specific challenge to the design of smart home systems [7].

Yet, the home context could specifically benefit from smart home solutions: inhabitants spend a large amount of time within their homes, and home activities have great potential to evolve and enhance themselves [6]. Smart home design could facilitate daily activities, support managing specific needs and domestic energy demand, meeting the challenges of digitalization, and enabling inhabitants' socialization [1, 6]. In this paper, we present a study that explores the home context through three user-driven steps. Viewing the home as an individual everyday environment we identify design directions to achieve a closer fit between design and context.

2 Related Work

To explore users' perspectives of smart homes, we will first look at the home context, what constitutes a 'smart home' and previous research engaging potential inhabitants.

2.1 The Home Context

The home is a complex place: It is inhabited by one or multiple users, who are subject to change. The technology they use, their relationships and responsibilities evolve over time – especially in multiple user homes, which are characterized by a shared ownership of tasks, individual levels of comfort, task completion and competing needs. Necessary negotiation over time [1] employs available surfaces for information sharing, such as fridge doors and wall displays [2]. Homes are built on specific values like emotional connection, self-expression and identity and therefore feature spaces of importance (e.g. the kitchen as the hub of home) [1]. Users' relationships with domestic technologies have been shown to focus on emotional attribution and attachment, which creates the feeling, or quality of home [3]. Home technologies have to recognize such emotional significances to avoid crossing "invisible value lines" and provide assistive tasks without challenging the identity of the user or their feeling of control [1].

Home behavior has been noted to be unstructured [4]. While that is largely true, domestic activities do feature routines revolving around particular timeframes, like morning routines or arriving-home behaviors. Once established, these routines are highly specific yet evolving procedures [5]. Events which cannot be routinized (e.g. activities with varying details), unexpected exceptions (e.g. illnesses) and other deviations can be experienced as stressful as well as positive: While a sudden sickness calls for improvisation and flexibility to accomplish necessary tasks [1], other behaviors only retain their meaning while not firmly embedded in a routine (e.g. spontaneous messages [2]). Accounting for routines as well as enabling deviations is most important to smart home systems [1, 5]. Routines can be difficult to automate [6] and rarely translate into predefined scenarios [1], but can be realized through customizable rules/scenarios [7].

2.2 Smart Homes

According to Balta-Ozkan et al. [8], a smart home is defined by four key aspects: a *communication network*, which connects *sensors* (and devices); *intelligent controls* for system management, and *smart features* that respond to user-, sensor- or system (data)

input. Devices in such a smart home can be controlled, accessed and monitored remotely. Consistent with this definition, many studies focus on features that make a home ‘smart’ [4, 8, 9] – for example by elucidating control preferences [10, 11], energy feedback preferences [12], energy monitoring [13], and acceptance/perceptions of smart homes [14–16]. In the following, we present relevant studies investigating user expectations and anticipations regarding smart homes and related technology.

A popular aspect is *home control*, i.e. the central control of appliances and devices. Previous studies identified different objects most important for users in different countries [10, 11, 14] – alarm and childcare systems, temperature and humidity control, smart cleaning, energy management, lights and TV being mentioned most, followed by blinds, plugs, switches as well as remote (smart phone) access.

Frequently, smart home visions focus on *energy feedback and energy management* [15] – placing the purpose and potential of the smart home within the realm of energy efficiency and employing a deeply instrumental view of smart homes [17]. Consumption (feedback) is a highly value-laden subject, covering issues of trust, privacy, environmental concern, and financial savings: In a study conducted in the UK, Germany, and Italy, participants expressed a lack of trust in power companies as well as concern about privacy and security when using a smart home system [15, 16]. Only 10% of German respondents would trust their energy provider with their data [11]. Interest in energy management has been shown to be based on environmental concern and/or on the wish to reduce costs [12, 16]. UK participants noted an increased agency through knowledge, while German and Italian participants valued transparent information about energy use most [16]. Feedback preferences from Finland show interest in consumption breakdown to individual appliances and comparisons with one’s own prior consumption [12]. UK households equipped with energy monitoring systems show a wide variation in their attempts to save energy, depending on whether and how the behaviors themselves were situated within their physical and social context, and whether saving behavior was ingrained before and independent of any use of energy monitoring systems [13]. Energy consumption feedback can also be coupled with time-variable tariffs, which are generally promoted as aiding energy conservation and reducing costs but are often met with reservations, mainly focused on low potential savings, practicality, and the difficulty of behavior change within appointed slots [12, 16, 18].

Communication features in smart homes have largely been limited to machine-human communication (e.g. notifications about [system] events), even in cases where users were involved in developing new services [5]. As argued earlier, the home is a place of social interaction and negotiation among humans. Features supporting such activities, have, however, been rarely discussed in previous studies.

Regarding user’s *expectations* of smart homes, German participants expected increased comfort, sustainability and security, lower heating and energy costs, and fun, though 29% of respondents could not name any reasons for using a smart home [11]. UK participants saw potential benefits in increased quality of life and leisure time, support of energy conservation, energy savings (reduced costs, environmental benefit), and assisted living [15]. They were skeptical, whether substantial savings could presently be achieved. UK and Italian participants also noted support for assisted living [16].

Users noted several *challenges* that currently prevent mainstream adoption of smart homes. Chinese and European participants noted privacy, data security, cost, reliability, and reliance as major barriers [11, 14, 15], as well as concerns about smart home technology being divisive and exclusive [15]. Further challenges included a lack of perceived benefits and fear of change [11] as well as high perceived complexity and unfamiliarity with smart home technology [2]. Some of these issues, such as privacy and reliability, have been perceived as primarily technical issues and not been addressed from a user’s perspective, setting the main design challenges in user-home interaction in the realm of security, privacy and trust, usability and user-friendly design [17].

Previous work shows a well-developed focus on investigating specific features yet lacking others as well as an overall view. While it therefore elucidates features and topics to base our research on, it also shows that “[...] a clear user-centric vision of smart homes is currently missing from a field being overwhelmingly ‘pushed’ by technology developers” [17, p. 464]. This paper therefore aims at providing design directions that include the everyday character and *context* of homes [1, 2] by combining expert and participant input in three phases. By that, we respond to user needs within the smart home [5] as well as address the challenges for smart home design [17, 19] outlined above, enriching but also moving beyond the research of relevant features.

3 Methods, Procedure and Materials

Our approach was divided in three main phases, each of which combined field-specific, scientific expertise and participants’ input (Fig. 1). It started with open self-observations of everyday interactions and the home context, and then moved via design scenarios to detailed design concepts. Each phase yielded results in the shape of features and system expectations, which were then analyzed as to their design implications.

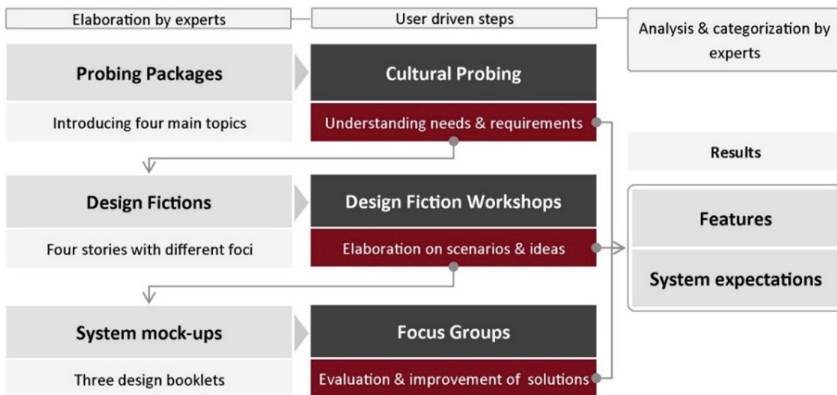


Fig. 1. The overall process: light colored boxes (left, right) indicate input and elaboration by experts (researchers, designers), while dark colored boxes (middle) indicate user involvement.

The procedure for each phase is described in the sections below, addressing the following aspects:

1. Deepen the understanding of everyday life interactions through an initial screening questionnaire, cultural probing and follow-up interviews. The aim was to elucidate user requirements, basic home principles and specific context factors.
2. Develop smart home design fictions using the findings of the cultural probing phase and expounding on (alternative) features with users in “design fiction workshops”.
3. Design and evaluate smart home components through focus groups which discussed and evaluated prototypical design solutions created from the workshop results.

3.1 Deepening the Understanding of Everyday Life Interactions

The first phase focused on understanding the practices of people in their everyday environment including routines in which they (potentially) interact with home technologies. The goal was to identify which design possibilities and opportunities could support and enrich these interactions. The main method applied was cultural probing [20], an in-depth measure typically using a small, carefully chosen sample. We employed online screening and theoretical sampling to achieve a well-balanced user group:

The **online screening questionnaire** elucidated different aspects of participants’ lives, including living situation, experience with home technologies, as well as general attitudes and interests. We employed an external user panel and received 251 completed questionnaires, of which 213 expressed a willingness to participate in the cultural probing study. The following set of criteria was used for recruiting ten participants:

- Achieve gender balance, diversity in age, educational background and tech. affinity
- Include different types of households (e.g. individuals with and without children)
- Include participants with different experience levels with relevant technologies, including both positive and negative experiences with existing appliances

While most participant characteristics could be reasonably controlled (Table 1), most screening respondents were technologically and environmentally interested and active. These skews are therefore also present within the cultural probing sample.

Cultural Probing is a method to gather data about people’s lives, values and thoughts to inspire and guide a subsequent design process [20]. *Probes* typically consist of artifacts (e.g. maps, postcards, camera) that are used by participants to record feelings and interactions related to a design problem. The main goal is to gather unsupervised responses over time. Our probe included a notebook, 20 action cards, materials to draw a map of their home and a single-use camera or photo-upload link. The diary-styled notebook contained prompts to document appliance use daily as well as eight photo tasks (e.g. “Please photograph technology that saves you time.”). The action cards focused on four home-related aspects, based on previous research (Sect. 2): energy use, home control, information, and communication and participation. Participants were asked to use two cards per day. Each included an action (e.g. “Please find appliances on stand-by and mark them on the map with the stickers provided.”)

Table 1. Cultural probing participants

Gender		Age group		Interest and Commitment	very – more – less – not
Male	5	19-35	3	Technology	6 – 3 – 1 – 0
Female	5	36-50	4	Environment	4 – 3 – 2 – 1
		51+	3	Community	4 – 1 – 2 – 3
Education				Politics	3 – 3 – 1 – 3
Compulsory education			1	Living situation	
Apprenticeship			3	Single	2
Vocational school			1	Domestic partnership, children	3
Secondary school			3	Domestic partnership, no children	3
College/University			2	Single parent	2

combined with an open-ended statement (“This appliance is usually on standby, because...”). The participants recorded their experiences for ten days, paying attention to different elements of their home-life. After a first analysis of their probe, each participant was invited for a follow-up interview, which first discussed their daily routines. Participants were asked why and when they use certain devices and how they create comfort and wellbeing. We then discussed each probing item (e.g. photo, card), with in-depth questions for items flagged as interesting or unclear in the analysis of the probe.

3.2 Development and Elaboration of Design Fictions

Investigating scenarios of smart home interactions helps understand the consequences and chances of such approaches. This phase of our research process therefore developed several fictional scenarios based on phase one, elaborating the fictional stories with potential users using the method of participatory design fiction [21].

Design fictions are a creative way to elaborate on speculative, but realistic visions of the future [21, 22]. They easily translate into design – originally conceived as a tool for designers, they have been used in a participatory manner with users [18]. Design fictions are less task-oriented and wider in scope than more commonly used user scenarios [23]. Being less focused on a specific technology, they shift the perspective to its implications. Placing the fiction in the future or in an alternative world helps to make this shift. In our study, experts developed four design fictions: written stories describing a day in the life of four fictional characters living with slightly different smart home systems. All four systems included the following aspects taken from cultural probing:

- Energy awareness and use, e.g. energy feedback, comparison, advice, master switch
- Home Automation, e.g. data access, automated control of home parameters
- Communication, e.g. messaging system, (virtual) bulletin board, community events
- Organization and Information, e.g. synchronizing calendar data, notifications

The design fictions were used in a series of four workshops with a total of 34 participants (Table 2). The workshops drew on two “Innovation Games” [24] to facilitate reflection and discussion. In the first game, participants formed groups of two to four members. Each group selected one of the four fictions and performed a SWOT-type analysis, identifying strengths, weaknesses, opportunities and concerns (renamed from “threats”), subsequently discussing identified aspects with the other groups. After collecting all aspects, each person distributed five points among the collected aspects to indicate their relevance. The groups then reformed and selected four aspects to elaborate on the consequences these aspects could have on people’s lives and smart home design. The second innovation game “Product Boxes” aims at generating key characteristics a future product should have. The groups were asked to imagine an ideal smart home system which would implement, extend or avoid aspects discussed earlier and design its packaging using materials like cardboard boxes, as if it was readily available in a store. Through this, participants did not only collect features but also imagined how the qualities of the product would be advertised. The groups then used sandwich panels and other crafting material to create physical mockups of their system (Fig. 2).

Table 2. Design fiction participants

Gender		Age group		Interest and Commitment				
				very	more	less	not	
Male	16	19-35	6	Technology	22	5	5	2
		36-50	9	Environment	24	8	2	0
Female	18	51-64	6	Community	12	14	8	0
		65+	3	Politics	6	11	11	6
Education				Living situation				
		Compulsory education	3	Single	12			
		Apprenticeship	6	Domestic partnership, children	6			
		Vocational school	6	Domestic partnership, no children	10			
		Secondary school	9	Flat share	5			
		College/University	10	Other	1			



Fig. 2. Example product boxes and system mock-ups from design fiction workshops.

3.3 Design and Evaluation of Smart Home Components

The third phase translated the high-level design fictions into tangible design solutions, evaluated and improved them involving potential users. In a first step, the results of the design fiction workshops – system characteristics, features, main concerns and mitigating factors – were analyzed and their key aspects identified. A professional user interface designer created interface mock-ups for each of the emerging aspects to serve as a discussion basis in focus groups. The individual designs (Fig. 3) did not form a complete system, but were combined into three booklets according to their main theme:

- *Communication & Information* included a digital neighborhood blackboard, a discussion forum, a personal messaging service, a petition support platform, a public display, a local newsletter, and a real-time public transport departure monitor.
- *Control & Automation* comprised of (automatically) controlling various appliances, e.g. lights, machine learning from user behavior, remote access and notifications.
- *Energy Feedback & Organization* included energy consumption feedback and notifications, a flexible tariff, a digital shopping list and a household calendar.

We hosted four focus groups with a total of 26 participants (Table 3). The focus groups consisted of a walk-through of all three booklets. Participants first browsed a booklet, annotating it individually. We then opened the discussion, using a set of guiding questions per screen, and letting the discussion develop before moving on to the next booklet. The questions were based on the results of previous phases and covered aspects of type of use (e.g. “In which situations do you want manual control?”), information needs (e.g. “What type of energy data do you want to see here?”), willingness to share or use (“Would you be willing to share this with others?”), expected benefits (e.g. “Do you think this would save you time or make your daily life more comfortable?”) and potential concerns (“What do you think happens if the system breaks down?”).

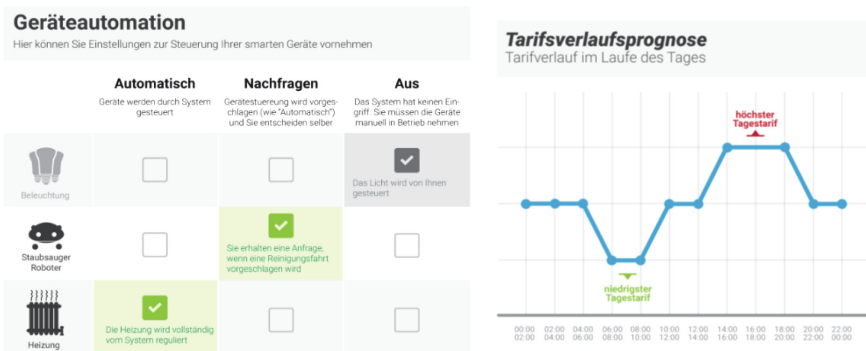


Fig. 3. Two example design mock-ups: the upper shows an aspect of home control - settings for lights, vacuum-cleaner and heating, and the lower information about the current tariff.

Table 3. Focus groups participants

Gender		Age group		Interest and Commitment					
				very	more	less	not		
Male	15	19-35	10	Technology	21	2	3	0	
		36-50	12	Environment	22	3	1	0	
Female	11	51+	3	Community	12	12	2	0	
		65+	1	Politics	6	7	7	6	
Education				Living situation					
		Apprenticeship		6		Single		8	
		Vocational school		3		Domestic partnership, children		4	
		Secondary school		10		Domestic partnership, no children		10	
		College/University		7		Other		4	

3.4 Analysis and Categorization

In the following, we give an overview of the data-streams resulting from the research phases, how they were categorized and synthesized.

Cultural probing yielded photos, diary entries, and interview statements, which were clustered to identify their main topic (e.g. routines). Being concerned with the status quo in homes, they were analyzed apart from the workshops and focus groups.

The **design fiction workshops** provided aspects identified and rated in the SWOT analysis, product boxes and system mockups, and discussion statements. SWOT topics were sorted, and scores consolidated across workshops. Product box and system mockup descriptions were separated into type of product (e.g. software, in-home monitor), control devices (e.g. dedicated remote), controlled appliances (e.g. heating, light), features (e.g. assistive functions), attributes (e.g. small, safe), anticipated consequences (e.g. growing lonely) and miscellaneous items (e.g. function-wise limitation of system internet access). Discussions in the **design fiction workshops** and **focus groups** were partitioned into individual statements. Each statement was coded according to the features, system abilities, mental models and expected consequences it contained. After compilation, each code (e.g. interest in remote control) was classified according to its main topic (e.g. home control) and category (e.g. expression of interest), see Fig. 4.

4 Findings: Desired Features and System Expectations

The main topics resulting from code analysis can be partitioned into two categories: *Features* (4.1) refers to seven specific areas of operation, or modules of the envisioned smart home system. *System expectations* (4.2) refers to more general system properties including expected consequences of system use and system abilities.

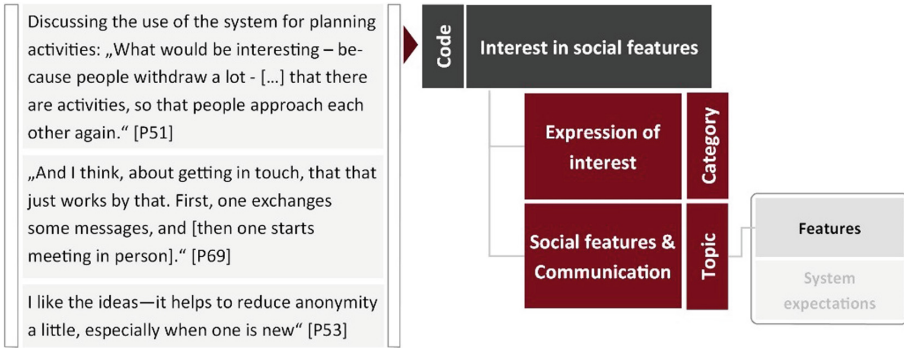


Fig. 4. An example of how statements (translated) were coded, categorized and clustered.

4.1 Features

Social Features and Communication was the most diverse feature including 27 different subtopics discussed on 5 categorical levels. It was among the top five rated SWOT topics and one of the most discussed topics in the focus groups (over 40 discussion instances). Participants were deeply interested in combatting anonymity (yet preserving privacy, 30 discussion instances), facilitating the initiation of contact and providing an easy means of communication within apartment buildings. It was most important to them that the features would not be designed to replace face-to-face contact (16 discussion instances) but enabling it. They expected that such features would increase community building and neighborly cooperation, assist in organizing activities or use of communal rooms, though some participants also voiced concern a potential misuse of such a “network” (e.g. surveillance, bullying, and presence-tracking).

Energy consumption monitoring was also a diverse subject, discussed on four categorical levels and 21 subtopics, including cost transparency and cost reduction, which were two of the top five rated topics in the SWOT analysis (7 points each). Additionally, energy feedback was most strongly spoken for (16 discussion instances voiced interest). In general, this interest was largely based on financial considerations (possible savings or rewards), personal attitudes such as energy saving as value, transparency as a need, sustainability as personal goal or economy/austerity as a moral good. Participants considered access to individual information as necessary to support or attain those goals. To them, the purpose of energy monitoring technology was increasing the intelligibility and transparency of energy use and costs, and empowering users to make informed decisions. Participants were interested in energy as well as water consumption. They noted the importance of visualizing costs as well as consumption units, a certain amount of accessible detail (breakdown to appliance-level or activities), and forecasting capabilities. Less popular options were normative comparisons, predict energy consumption with new devices and the ability to compare energy suppliers. Participants’ greatest concern was consumption monitoring being a time intense task and therefore only usable for limited time periods, preferring ambient information for continuous long-term.

Home Control included 15 different subtopics on two categorical levels (feature description and expression of interest). Remote control was the most discussed single feature by far (28 discussion instances) and among the top five rated topics in the SWOT analysis (7 points). Participants expressed a need of increased flexibility and convenience, which they would realize by adapting home control to individual routines and circumstances. Participants made a distinction of technology “embedded” in the home (e.g. lighting, heating) and appliances (e.g. washing machine, personal devices), connotated by a different emotional attachment, which nonetheless should both ideally be remotely controllable via app, SMS or web interface. Participants’ concerns included reduced accessibility for people without smartphones, data-tariffs, technological knowhow or capacity. Other barriers were perceived security issues, the unclear general setup (e.g. contract and legal conditions), and its use in multiple user homes. They mentioned the necessity of automation, appliance programming, and fine-grained control-models (e.g. rooms, appliances) to use smart home control in a real-life setting.

Information Features were already strongly incorporated in households taking part in the cultural probing. Participants employed different surfaces to share mobility information, information about the surrounding area, timetables and individual information. In a virtual environment, they would additionally store household information (e.g. device manuals, shopping support e.g. light bulb types). Participants noted that sharing regional information encourages community building and feelings of belonging.

Safety and Security contains features for the living space (seven subtopics) and for assisted living (four subtopics). The former included alerts of trouble within the home (e.g. fire, water damage) and other features that would make participants feel safer, like burglar deterrents (light automation and remote access), physical access control and door/window-sensors. The latter discussed features like assistive home control and household monitoring by family members. Their feelings of security were heightened by features concerned with support in emergency situations as well as by the general overview of one’s household, appliance use and energy consumption.

Tariff information and transparency were expressed as important features, especially with a time-variable tariff. Participants equated a variable tariff with a time-of-use tariff (i.e. fixed time slots of different pricing), which was seen as both potentially positive and negative – as an opportunity for transferring “favorable price-realities” to consumers as well as for increasing overall costs and controlling individual energy consumption by energy providers. A need for comfort, which many participants were not willing to compromise within their own home, was a major barrier. Participants noted that the potential to shift activities would be small, and the time and effort in managing consumption would largely exceed the savings, which were perceived negligible. They did note that shifting potential could be increased if an energy monitoring system allowed remote access to appliances and automating domestic actions. Another mitigating possibility were temporary energy storage capacities (e.g. batteries) in the home.

Participation Features. Within the cultural probes and succeeding interviews, participants described that they frequently notice problems in their immediate environment and have ideas for improvement. While some participants stated they have satisfactory possibilities to become active, others felt dialog was currently not properly

facilitated to hear their complaints or realize changes. Smart home features that enable participating, contributing to initiatives as well as simple reporting for problems encountered in the building or surrounding area were met with interest by these participants.

4.2 System Expectations

Customizability was a main expectation: tailoring features (4 SWOT points, 9 discussion instances) and modularly creating systems mitigated many perceived privacy issues and increased the perceived usefulness of a smart home. Participants' choices were strongly related to their individual level of comfort with technology and need of control within the home. Within all phases, participants expressed divergent goals – e.g. a wish to gain increased awareness and transparency as well as support for different values (e.g. sustainability, economy, privacy). They therefore perceived a system as valuable only if it specifically catered to their needs and accommodated these personal goals.

Automation and intelligent learning algorithms were a prerequisite to realize true flexibility in participants' minds – especially if a system includes time variable tariffs. On the other hand, others noted that automation and learning behavior could disempower users, creating a communication gap between them and their homes and thereby reducing home comfort as well as increasing surveillance and individual “transparency”.

Accessibility. Imagining their ideal systems, participants expressed a need for low complexity, which would make it accessible for varying user groups. Systems should cater to people with less technical knowledge and experience as well as people with disabilities by including specific, user group-oriented designs and support. Participants also discussed secondary users, such as persons visiting equipped households (e.g. grandparents babysitting their grandchildren) that need to be able to interact with the system.

Compatibility with existing devices and home appliances was another factor for the acceptance and attractiveness of a smart home – the necessity to refit an entire household in introducing such a system was a great barrier, and possibilities for extension to other areas like car sharing and other applications already in use highly desirable.

Data Use, Privacy and Unauthorized Access. Participants had little trust in data protection measures provided by energy providers or connected facilitators. They were therefore skeptical to connect their (imagined) home to the Internet or other networks. While some saw this as unavoidable and resigned themselves to the cost, “paying” for services with their data and privacy, others called for settings allowing detailed control of data-transfer, appropriate data encryption and the localized storage of (critical) data. Concerns related to privacy loss included surveillance by energy providers and other third parties, but also monitoring by other household members. That their daily actions within their own homes might become traceable was unimaginable to some participants and irreconcilable with their needs. This traceability and possible data misuse received

eight SWOT points each, while the “transparent customer” received 10 SWOT points. The concern of unauthorized access acted as further barrier to remote control features.

Reliability was met with skepticism. Participants did not perceive current technologies as stable and usable enough to entrust them with their homes, and insisted on local and manual control as fail-safe if the system is unresponsive or connection is lost. They questioned the (hidden and uncommunicated) costs of support, and noted that liabilities might not always be clear (e.g. what would happen in the case of data misuse).

Time expenditure and long-term use were both framed negatively: Participants assumed that system use would generally lead to a higher effort to manage their lives, especially with time-variable tariffs, rating excessive planning of daily lives as one of their top five concerns (2 SWOT-points and 5 discussion instances). This high system cost (personal energy, money, time) was contrasted with comparably small benefits over time, also noting several negative consequences of long term use. Of these, dependency on the system and on technology, “dumbing down through technology use” and loss of independence were the most important subjects (12, 8 and 3 SWOT points, 2, 8 and 10 discussion instances respectively). Yet, they also noted that high system costs were based on bad design, namely the overall poor usability of home technology and technology-centrism of systems, and that clear, sustained benefits could make time-intense processes more attractive (e.g. financial incentives).

5 Design Implications

Based on these results we developed recommendations to inform the design of future smart home solutions, structured into *design dimensions* and *design goals*:

5.1 Design Dimensions

Design dimensions address background factors designers should consider to complement the more system-centric “user context” [19]: Dimensions apparent in this study, in line with [1, 2], are time, space, relations, individual factors and values.

Time means designing in the context of activities, routines and household evolution as well as dealing with exceptions and change. To design along the dimension of time means asking whether a design allows for temporal expression of the home context, and how time is (re-)produced in a system (e.g. schedules). *Space* addresses designing for spaces with specific functions and meanings. To design along the dimension of space, designers should ask how a system interacts and integrates with spatial dimensions of the home, where it is located and whether it creates new locations or spatial links. To design along the dimension of *relations*, design must acknowledge multiple users, shared ownerships within the home, and negotiation of competing needs and varying levels of comfort. A system designed for relations must additionally account for social practices and relationships constraining as well as enabling actions (e.g. primary responsibility for domestic labor) [13]. *Individual factors* focus on additional individual characteristics and abilities (e.g. computer literacy). Systems can be viewed by the in- or exclusion of factors, and thereby, of people, in their design.

Designing along these dimensions therefore also addresses users' concerns about smart technologies excluding specific groups (this study, [15]). *Values* address feelings, motivations, views and goals and frame the actions and activities in the home context e.g. being economic or environmentally concerned (this study, [12, 16]). Designing along this dimension therefore includes consideration which values, motivations and actions a system chooses to promote and how they relate to pre-existing value structures.

While these design dimensions can be viewed separately, they are in fact intersectional: for example, negotiation was important for all non-single households, combining the use of specific spaces for information storage (*space*) and the practice of sharing (*relations*). Similarly, the importance of routines (*time*) can be seen in users fearing apathy or "losing" household routines (*values*). Thus, design along one dimension should acknowledge its effects on another and take its prerequisites into account.

5.2 Design Goals

In contrast to design dimensions, design goals provide aspects design should aim for. Based on the collected material we derived seven design goals:

Designing for control means control of lives rather than control of systems, encompassing felt agency, system reliability as well as control of data use and privacy. Privacy and data security were often framed as issues of choice and control. Participants saw smart home systems as vehicles for increased agency through knowledge and transparent information, yet also anticipated loss-of-control feelings (e.g. "over-automation") and actual loss of control (e.g. system malfunction). This mirrors privacy concerns across China and Europe [11, 14, 15]: Storage and eventual use of personal information (selling, use in advertisement) were as much issues as monitoring with third parties potentially knowing daily routines and home occupancy.

Designing for low effort focusses on low time expenditure and practicability. This can be achieved through various means, including system control and user-system communication. Of these, automation can be one, though barriers in the form of technological immaturity, low intelligibility and impractical user control persist. Low effort has been especially discussed in connection to time-variable tariffs: Similar to previous studies our participants were hesitant about their use and practicability [12, 16]).

Design for integration with everyday life practices takes a slightly different angle on design for control and design for low effort: People spend a substantial amount of their time at home, engaging in different activities and practices. Everyday activities are the focal point of what it means to be 'at home'. Integration with everyday life practices thus focuses on integrating smart functionalities with everyday practices and routines, without interrupting process flows or forcing users to shift their focus from the tasks at hand towards details of system control. Smart functionalities should offer additional possibilities that support users in achieving their goals, but not require them to modify their behavior. They should furthermore aim to minimize gaps and discrepancies in media usage (e.g. the need to switch between devices), physical location (e.g. having to

go to another room to access the control point) or atmospheric dispositions (e.g. need to interact with a very demure control interface in an entertainment context).

Designing for evolvability includes learning as well as maintained compatibility. While designing for initial compatibility is an important system component and architectural consideration, maintained compatibility is integral for system evolution. From a system's point of view, evolvability can be noted as one component of a system's agility [19]. Only evolvable systems are able to reflect changing household needs and changed circumstances such as e.g. children entering school.

Designing for identity is especially relevant for the home context: Home fulfills important functions in enabling and expressing identity besides more mundane function of providing shelter and living space. People use their home to express and 'materialize' their identity. Smart home concepts should support these practices and different ways of self-definition (including non-technical ones), especially in the view of enabling appropriation and sustainable smart home development.

Designing for sociability including communication, cooperation, and participation, should be one of the main design goals in system creation to support sociability and participation as values: Active decision making, partaking of communal development and communication proved important personal motivators for our participants in the specific context of an urban environment. It remains to be investigated if this topic is of importance in different settings and how well technology might integrate with "value lines" in the social context. We also note that interest in communication, activity and participation have strong social desirability, possibly biasing contributions to this topic.

Design for Benefits. In the smart home context, benefits include added comfort, increased convenience, security and flexibility. Benefits need to compensate for costs – in this study as well as in [15] these related to installation, repairs and maintenance, vulnerability to rising prices and non-monetary costs (data and privacy as currencies). Yet, design for benefits should generally aim higher than mere acceptance of a system or counterbalancing perceived and actual costs. It should address users' perceived lack of benefit [11, 15, this study] and specific beneficial features punctually, but needs to include real, tangible, and significant benefits to user's lives *by design*. Features beyond the traditional realm of smart homes, in particular social, informative and participative features potentially increased the value of systems for participants in this study.

6 Conclusions

This study describes one of currently few attempts to integrate contextual user requirements analysis with application-oriented design thinking. Such an integrated approach is necessary to leverage the advancement of the 'Smart Home', which has so far been driven more strongly by technology and business push than by genuine user demand. Viewing the home as an individual everyday environment has enabled us to identify design directions that achieve a closer fit between design and context. Based on extensive empirical data we identified important features and system expectations relevant to users. From this, we derived design dimensions and goals that can help

developers to understand the context of use and provide decision support for the prioritization of design directions.

The study further contributes with a methodological concept that encompasses three successive steps: cultural probing, participatory design fiction and focus groups. To our experience, this specific qualitative research process fulfilled the goal to progress from an understanding of the home context “as-is” towards a blending of requirements with technological solutions. Future research should investigate optimizing the effort and execution timeframe of such combined qualitative approaches, while still maintaining the richness of contextual design insights. This would make this approach even more applicable, especially for smaller projects in commercial development settings.

Our findings provide a comprehensible framework for the design landscape for smart home systems and serve as a basis for integrating previous findings into an overall framework. Subsequent research should explicitly build on the design dimensions and specific design goals we derived. The focus for such further investigations should be to assess their actual value in a design process, in order to develop them further with regard to completeness and relevance for implementation work.

Acknowledgements. This research was performed within the Smart City Demo Aspern (SCDA) project (financed by the Climate and Energy Fund of the Austrian Research Promotion Agency) and the SIM4BLOCKS project (funded from the European Union’s Horizon 2020 research innovation program under grant agreement No. 695965).

References

1. Davidoff, S., Lee, M.K., Yiu, C., Zimmerman, J., Dey, A.K.: Principles of smart home control. In: Dourish, P., Friday, A. (eds.) *UbiComp 2006*. LNCS, vol. 4206, pp. 19–34. Springer, Heidelberg (2006). https://doi.org/10.1007/11853565_2
2. Taylor, A.S., Harper, R., Swan, L., Izadi, S., Sellen, A., Perry, M.: Homes that make us smart. *Pers. Ubiquitous Comput.* **11**, 383–393 (2007). <https://doi.org/10.1007/s00779-006-0076-5>
3. Mäyrä, F., et al.: Probing a proactive home: challenges in researching and designing everyday smart environments. *Hum. Technol. Interdiscip. J. Hum. ICT Environ.* **2**, 158–186 (2006)
4. De Silva, L.C., Morikawa, C., Petra, I.M.: State of the art of smart homes. *Eng. Appl. Artif. Intell.* **25**, 1313–1321 (2012). <https://doi.org/10.1016/j.engappai.2012.05.002>
5. Coutaz, J., Fontaine, E., Mandran, N., Demeure, A.: DisQo: a user needs analysis method for smart home. In: *Proceedings of the 6th Nordic Conference on Human-Computer Interaction Extending Boundaries, NordiCHI 2010*, pp. 615–618. ACM, New York (2010)
6. Davidoff, S., Zimmerman, J., Dey, A.K.: How routine learners can support family coordination. In: *Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010*, p. 2461. ACM Press, New York (2010)
7. Woo, J., Lim, Y.: User experience in do-it-yourself-style smart homes. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp 2015*, pp. 779–790 (2015)
8. Balta-Ozkan, N., Davidson, R., Bicket, M., Whitmarsh, L.: The development of smart homes market in the UK. *Energy* **60**, 361–372 (2013). <https://doi.org/10.1016/j.energy.2013.08.004>

9. Koskela, T., Väänänen-Vainio-mattila, K.: Evolution towards smart home environments: empirical evaluation of three user interfaces. *Pers. Ubiquitous Comput.* **8**, 234–240 (2004). <https://doi.org/10.1007/s00779-004-0283-x>
10. Kühnel, C., Westermann, T., Hemmert, F., Kratz, S., Müller, A., Möller, S.: I'm home: defining and evaluating a gesture set for smart-home control. *Int. J. Hum Comput Stud.* **69**, 693–704 (2011). <https://doi.org/10.1016/j.ijhcs.2011.04.005>
11. Deloitte Consulting GmbH: Technische Universität München: Ready for takeoff ? Smart home aus Konsumentensicht (2015)
12. Karjalainen, S.: Consumer preferences for feedback on household electricity consumption. *Energy Build.* **43**, 458–467 (2011). <https://doi.org/10.1016/j.enbuild.2010.10.010>
13. Murtagh, N., Gatersleben, B., Uzzell, D.: 20:60:20-differences in energy behaviour and conservation between and within households with electricity monitors. *PLoS ONE* **9**, e92019 (2014). <https://doi.org/10.1371/journal.pone.0092019>
14. Zhai, Y., Liu, Y., Yang, M., Long, F., Virkki, J.: A survey study of the usefulness and concerns about smart home applications from the human perspective. *Open J. Soc. Sci.* **02**, 119–126 (2014). <https://doi.org/10.4236/jss.2014.211017>
15. Balta-Ozkan, N., Davidson, R., Bicket, M., Whitmarsh, L.: Social barriers to the adoption of smart homes. *Energy Policy* **63**, 363–374 (2013). <https://doi.org/10.1016/j.enpol.2013.08.043>
16. Balta-Ozkan, N., Amerighi, O., Boteler, B.: A comparison of consumer perceptions towards smart homes in the UK, Germany and Italy: reflections for policy and future research. *Technol. Anal. Strat. Manag.* **26**, 1176–1195 (2014). <https://doi.org/10.1080/09537325.2014.975788>
17. Wilson, C., Hargreaves, T., Hauxwell-Baldwin, R.: Smart homes and their users: a systematic analysis and key challenges. *Pers. Ubiquitous Comput.* **19**, 463–476 (2015). <https://doi.org/10.1007/s00779-014-0813-0>
18. Prost, S., Mattheiss, E., Tscheligi, M.: From awareness to empowerment : using design fiction to explore paths towards a sustainable energy future. In: *CSCW 2015*, pp. 1649–1658 (2015). <https://doi.org/10.1145/2675133.2675281>
19. Solaimani, S., Bouwman, H., Secomandi, F.: Critical design issues for the development of Smart Home technologies. *J. Des. Res.* **11**, 72–90 (2013). <https://doi.org/10.1504/JDR.2013.054067>
20. Gaver, B., Dunne, T., Pacenti, E.: Design: cultural probes. *Interactions* **6**, 21–29 (1999). <https://doi.org/10.1145/291224.291235>
21. Grand, S., Wiedmer, M.: Design fiction: a method toolbox for design research in a complex world, pp. 1–25. *Designresearchsociety.Org* (2006)
22. Sterling, B.: *Shaping Things*. MIT Press, Cambridge (2005)
23. Blythe, M.A., Wright, P.C.: Pastiche scenarios: fiction as a resource for user centred design. *Interact. Comput.* **18**, 1139–1164 (2006). <https://doi.org/10.1016/j.intcom.2006.02.001>
24. Hohmann, L.: *Innovation Games: Creating Breakthrough Products Through Collaborative Play*. Addison Wesley, Boston (2006)



A Clustering Approach for Profiling LoRaWAN IoT Devices

Jacopo Maria Valtorta^{1,3}, Alessio Martino¹ , Francesca Cuomo^{1,3} ,
and Domenico Garlisi^{2,3} 

¹ Department of Information Engineering, Electronics and Telecommunications,
University of Rome “La Sapienza”, Via Eudossiana 18, 00184 Rome, Italy

valtorta.1620736@studenti.uniroma1.it,
{alessio.martino, francesca.cuomo}@uniroma1.it

² University of Palermo, Palermo, Italy

domenico.garlisi@unipa.it

³ Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT),
Parma, Italy

Abstract. Internet of Things (IoT) devices are starting to play a predominant role in our everyday life. Application systems like Amazon Echo and Google Home allow IoT devices to answer human requests, or trigger some alarms and perform suitable actions. In this scenario, any data information, related device and human interaction are stored in databases and can be used for future analysis and improve the system functionality. Also, IoT information related to the network level (wireless or wired) may be stored in databases and can be processed to improve the technology operation and to detect network anomalies. Acquired data can be also used for profiling operation, in order to group devices according to their characteristics. LoRaWAN (Long Range Wide Area Network) is one of the emerging IoT technologies in today’s world, it is a protocol based on LoRa modulation. In this work, we propose a methodology to process LoRaWAN packets and perform profiling of the IoT devices. Specifically, we use the k -means algorithm to group devices according to their radio and network behaviour. We tested our approach on a real LoRaWAN network where the entire captured traffic is stored in a proprietary database. Our analysis, performed on 286,753 packets with 765 devices involved, leads to remarkable clustering performance according to validation indices such as the Silhouette and the Davies-Bouldin indices. Further, with the help of field-experts, we were able to analyze clusters’ contents, revealing results both in line with the current network behaviour and alerts on malfunctioning devices, remarking the reliability of the proposed approach.

Keywords: IoT · LoRa · LoRaWAN · Machine Learning · k -means · Anomaly detection · Cluster analysis

1 Introduction

The Internet of Things (IoT) is a new technology paradigm envisioned as a global network of machines and devices capable of interacting with each other. According to the IoT Analytics forecast of 2018 [12], the market for IoT has seen an unexpected acceleration in the first months of the 2018. Currently, the number of connected devices exceeds 17 billions, and the number of IoT devices is 7 billions. The focus of the IoT is to interconnect together things or smart devices in order to create smart environments. Each device is an appliance with embedded electronics and software which can work as a sensor or actuator. Sensors are able to collect the state of some metric like temperature or air quality of the environment. Actuators are responsible for changing the state of the environment, e.g. open the window in presence of bad air quality. For this purpose, IoT devices exchange data and, in most cases, data are stored and processed by a central server. Moreover, the collected data can be used to perform device analysis and, in most of cases, the results of the analysis are focused on system optimization. LoRaWAN is an emerging technology in the IoT world proposed by LoRa Alliance and based on LoRa modulation. LoRaWAN uses the ISM band and is capable to connect thousands of IoT devices or End-Devices (EDs) in a geographic area of square kilometers. LoRa modulation presents an high energy efficiency and EDs can remain active for several years before replacing the battery package. Packets sent by EDs are collect by GateWays (GWs) that are deployed in the covered geographic area. Packets are forwarded from the GWs to the Network Server (NS). NS is responsible to process the packets, forward related information to the IoT applications and store the collected data (see Fig. 1 for the reference network architecture).

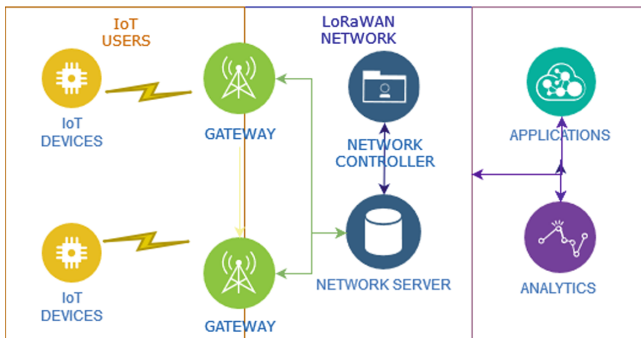


Fig. 1. LoRa network architecture

In this work, we propose and apply Machine Learning (ML) techniques to perform profiling of the IoT devices. We develop a framework that, starting from a database storing packets collected through the LoRaWAN technology, produces device profiling. This means that, by leveraging these ML tools, we are

able to derive profiles of the behaviour of IoT EDs connected to the LoRAWAN network. This tool has different applications:

- it allows to monitor the system behaviour and capture anomalies;
- it allows the network operator to use in an efficient way the system and to optimize the network planning;
- it allows to identify different *spaces* where service provision can be enhanced, that is, new radio resources, more suitable parameter settings, different configurations of the IoT devices and services.

For this work, we use the collected data of a real LoRaWAN network deployed in Italy. Starting from the LoRaWAN packet structure, we extract the relevant packet fields that characterize the behaviour of the device at physical and network layer. We use these fields to extract a set of features that represent the input of the ML algorithm. We apply an unsupervised learning approach to model the underlying structure or distribution in the data in order to learn more about the data. Specifically, the idea is to use the k -means algorithm to perform a grouping (cluster) analysis that identifies commonalities in the data. Alongside the application of the k -means algorithm, we perform the study of the best k value required as input of the k -means algorithm, namely the optimal number of clusters for the dataset at hand. Finally we study the peculiarity of the devices belonging to each of the resulting clusters. The main contribution of this work is two-fold. First, we identify how to apply a ML approach based on k -means to an IoT network and, second, we experiment this approach on a real LoRaWAN system. While in the recent literature there are papers dealing with the adoption of ML for IoT (see e.g. [4, 14, 23]), the literature lacks in the application of this methodology to LoRaWAN Networks. The remainder of the paper is organized as follows. Section 2 presents the main related works. Section 3 briefly recalls the LoRaWAN architecture and service. The ML approach is presented in Sect. 4 while the resulting profiling mechanism and the relevant analysis are discussed in Sect. 5. Finally, Sect. 6 concludes the paper.

2 Related Works

The interest in applying ML approaches to IoT systems is growing fast. Several papers proposed to use ML for anomaly detection or security issues [1, 31]. The paper by Bhatt et al. [4] focuses on the development of a hybrid network anomaly detection system that, by making use of ML techniques, is able to effectively detect malicious traffic data. Tailored to the dimensionality reduction in learning models induced for IoT networks, in [25] the authors showed that it is possible to induce highly accurate unsupervised learning models with reduced feature set sizes, which enables to decrease the required computational resources. Also new datasets are required for the development and testing of novel ML techniques. Indeed, in [24] the authors build a novel dataset from the wireless network packet traffic flow captured through Wireshark that holds different attack profiles. The profiling issues are also very interesting in the

IoT field since they pose new challenges and pave the way to new applications. From network data it is possible to extract users' behaviours as shown in the paper by Tao et al. [29] where they extract users' activities feature from inter-arrival time and packets' traffic size, and investigate their relationships. Network behaviour, as instead, can be recognized by applying finite state automata to IP flows records as shown in [9]. For example, Marchette [17] in 1999 proposed two interesting clustering methods applied to network data. These techniques allow the clustering of machines into "activity groups", which consist of machines which tend to have similar activity profiles. Here the first aim of the author is to apply these methods in security domains; in fact, they allow the user to determine whether current activity matches these profiles, and hence to determine whether there is "abnormal" activity on the network. Zhang et al. in [32], presented a k -means-based approach for clustering data packets in wireless multi-hop networks. They addressed the problems existing in such networks, namely imbalance of node power consumption and unfairness of node transmission, and addressed the trade-off between the energy consumption and other factors affecting the wireless multiple hop networks thanks to k -means clustering. A similar work based on IoT and Low-power Wide Area Networks has been done by Barrachina-Muñoz et al. in [3], where they proposed a simple reinforcement learning algorithm based on ϵ -greedy to enable reliable and low consumption Low Power WAN multi-hop topologies. Kim et al. [11] analyzed the transmission mechanism inside a LPWAN. The authors proposed a method which employs a k -means clustering algorithm to classify devices according to the traffic characteristic. Each cluster is assigned to a different priority in order to optimize channel access times: this, in turn, allows to avoid collisions and improve transmission efficiency. In the same year, Zhang et al. [33] studied an adaptive clustering algorithm for dynamic heterogeneous wireless sensor networks in order to adapt the dynamic change of topology in such networks. Their model dynamically elects cluster heads according to each node's energy and according to the average network energy, yielding longer network lifetime. Del Campo et al. [7] presented an LPWAN communication architecture for a real-time monitoring systems in power distribution grids, in the framework of the MAIGE project. MAIGE main goal is the experimental evaluation of innovative technologies for their massive deployment. The project describes a variety of use cases that demand different communication requisites. Mostafa, in [22], has given a more general and theoretical view on monitoring IoT network. To tackle the problem of monitoring the network and leave it unconstrained during its normal operation, he proposed several integrated graph-based optimization models and efficient algorithms for monitor placement and scheduling problems. For what concerns security issues, Kumar et al. in [13] analyze devices traffic in order to detect malware activities. They present EDIMA, a distributed modular solution which can be used towards the detection of IoT malware activity in large-scale networks by means of supervised ML algorithms. While clustering algorithms have been widely used to preserve security in the IoT devices, to classify devices, to adapt the dynamic change of network topology or, in general,

to analyze demands on the traffic characteristics, a detailed study on profiling IoT LoRaWAN devices and their traffic features in order to improve the network performance is still not available in the current literature.

3 LoRa Technologies and LoRaWAN Protocol

This section provides an overview of the LoRaWAN technology, at physical and network level. First, we will provide a presentation of the modulation scheme, afterwards, we will present the packet structure of the LoRaWAN, and how the received packets are stored in a real system. This real system is a deployment database of an IoT operator present in Italy.

3.1 LoRa Modulation Scheme

LoRa is a new long-range communication technology proposed by Semtech [27] few years ago. LoRa is a proprietary technology and is based on LoRa modulation. LoRa implements a chirp spread spectrum modulation that uses the entire wideband linear frequency to modulate chirp pulses. A chirp is a sinusoidal signal whose frequency increases or decreases over time, that encodes a certain number of information bits. Conversely to the most common FSK modulation, LoRa modulation maintains the same low-power characteristics, but improves the noise and interference immunity and, consequently, increases the communication range. The result is that a single GW can cover a region of different square kilometers.

While LoRa defines the physical layer and is a proprietary technology, LoRaWAN specification defines the network layer, this specification is publicly available and it is promoted by the open-source LoRa Alliance [28]. As shown in Fig. 1, the LoRaWAN protocol stack is based on three main components:

1. ED: is the low-power consumption sensor that communicates with gateways using LoRa modulation;
2. GW: is the intermediate element that forwards packets coming from EDs to a NS over an IP backhaul interface, such as Ethernet or 3G. There can be multiple GWs in a LoRa deployment.
3. NS: is the server responsible for deduplicating and decoding the packets sent by the devices. Afterwards, the related packet information is sent to the application. The NS can be also generating packets that should be sent back to the EDs, when ED configuration is required or in case of device-to-device communication (these are sling-shot through the network server).

The LoRaWAN network has a star-of-stars topology and, differently from traditional cellular networks, the EDs are not associated with a specific gateway. LoRaWAN does not enable device-to-device communications, packets can only be transmitted from an ED to the NS, or viceversa. NS provides network information to the network controller and to the analytic module, when they are

present. The network controller uses the data to perform a network optimization by tuning the EDs. The analytic module is an high-level application that presents statistical information about the network to the network operator.

In LoRa, EDs support multi-rate by exploiting six different Spreading Factors (SF), from 7 to 12. The selection of the SF has an impact on duration and delivery probability of the generated packet. Communication on different SFs in the same channel are in principle orthogonal [5]. In LoRa, basic chirps are simply a ramp from f_{min} to f_{max} (up-chirp) or from f_{max} to f_{min} (down-chirp). Chirps are cyclically-shifted to produce different symbols, and this cyclical shift carries the information. A symbol, with a length of N chips, can be cyclically shifted from 0 to $N - 1$ positions. The reference position is given by the un-shifted symbols at the beginning of the LoRa packet, present in the packet preamble. The SF defines two fundamental values: (i) the number of chips contained in each symbol is $N = 2^{SF}$; (ii) the number of raw bits that can be encoded by that symbol is SF; The LoRa Data Rate (DR) depends on the Bandwidth (BW) in Hz, the SF and the Coding Rate (CR) as:

$$DR = SF \cdot \frac{BW}{2^{SF}} \cdot CR \quad (1)$$

where the symbols/s are given by $BW/2^{SF}$ and the channel coding rate CR is $4/(4 + RDD)$ with the number of redundancy bits (RDD) from 1 to 4 used for the cyclic redundancy check (CRC). The adopted bandwidth can be configured as well: 125 kHz, 250 kHz and 500 kHz (typically 125 kHz for the 868 ISM band). The combination of an high SF and a small bandwidth produces a more robust transmitted signal that can cover very large distances (more than 10 km). LoRaWAN specification also provides an Adaptive Data Rate (ADR), an algorithm to estimate the best SF and transmission power values for each ED according to the Signal to Noise Ratio (SNR) perceived by GW: typically, the ADR is implemented in the network controller. Regarding the message type, the LoRaWAN protocol establishes that a Confirmed-data message must be acknowledged by the receiver, whereas an Unconfirmed-data message does not require an acknowledgment. During this study, we faced with the majority of messages of Unconfirmed-data type. In LoRaWAN, the system capacity is larger because the receiver can detect multiple simultaneous transmissions by exploiting the orthogonality when different SFs are used. Moreover, if the multiple simultaneous transmissions are generated with the same SF, a low difference in the signal strength (few dB values) can generate a channel capture effect that ensures the correct reception of the stronger signal. These features enable a LoRaWAN network to have a very high capacity and make the network scalable. A network can be deployed with a minimal amount of infrastructure and, as larger capacity is needed, more GWs can be added. Other LPWAN alternatives do not have the scalability of LoRaWAN due to technology trade-offs. In LoRaWAN, MAC commands can be used from the NS to configure ED parameters such as SF or power transmission.

3.2 LoRaWAN Network and Packet Structure

The LoRaWAN terminology distinguishes between uplink and downlink messages. EDs send uplink messages to the NS. Downlink messages are sent by NS to only one ED and are relayed by a single GW: they usually contain MAC commands, useful to customize the parameters used for the communication between the ED and the network. LoRaWAN messages used for the radio physical layer have the same format both for uplink and downlink. As shown in Fig. 2, at physical layer (the top of the figure) the LoRa packet is composed by the preamble, the physical header (PHDR), the physical header cyclic redundancy check (PHDR_CRC), the physical payload (PHYPayload) and the CRC of the packet. The PHDR is mandatory both for uplink and downlink messages, while the CRC is mandatory only in the uplink. The PHYPayload carries the MACpayload, the MACheader and the cryptographic message integrity (MIC). The MACHeader contains information about the LoRaWAN version used (v1 or v2) and the Message Type (MType). The Mtype field enables to distinguish registration packets (Join-Request/Accept) from Unconfirmed-data and Confirmed-data packets. MIC is a code computed over the MHDR. The MACpayload contains Frame Header (FHDR), Port Field (FPort) and the Frame Payload (FRMPayload). As for the MHDR, the FHDR field has information about the ED short address (DevAddress) as well as other control information carried in the Frame Control field (FCtrl) such as the status of the ADR for the communication. The Frame Port (FPort) field has a 0 value in case of FRMPayload containing only MAC commands while it is used by the application to discriminate the content of the payload, so the value of the packet is application-specific. FRMPayload is the payload containing MAC commands or application data, which is encrypted using AES with a key length of 128 bits. Note that, together with the physical layer messages, the NS also receives additional information regarding the

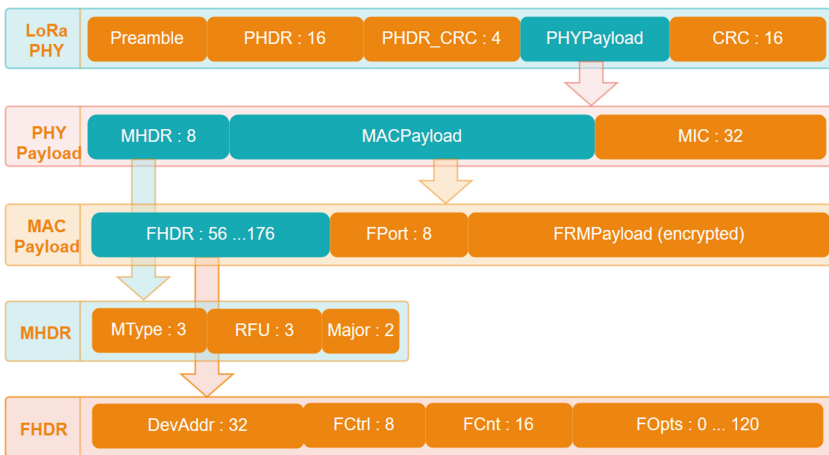


Fig. 2. LoRaWAN packet structure [bit]

physical parameters of the communication, such as SNR and Received Signal Strength Indicator (RSSI). Each ED has a packet counter (FCnt field) to keep track of the amount of data packets sent to the NS. The DevEUI is a global ED ID in IEEE EUI64 address space that uniquely identifies the ED, while the DevAddr consists of 32 bits address and identifies the ED within the current network (the DevAddr is allocated by the NS of the ED once it joined the network successfully). This packet format represents the starting point for the database structure that contains all the received packets.

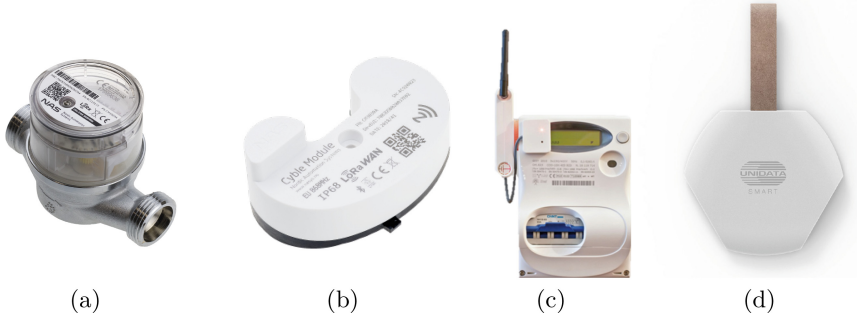


Fig. 3. Four different types of EDs used in the analysed network. (a) water meter; (b) gas meter; (c) energy meter; (d) GPS tracker.

For this work, we used a database of a real LoRaWAN network deployment by an Italian operator. The goal of this IoT national network is provide several application services, mainly related to the metering operations. The main application services provided by the network are: (1) water metering; (2) gas metering; (3) energy consumption metering; (4) GPS tracking; (5) smart road light; Fig. 3 presents 4 different devices used in the LoRaWAN network, and detailed in the figure caption. The total number of installed devices is 765, while the number of GWs is 49, they are disseminated on the whole Italian territory, though the 80% of them are deployed between Milan and Naples. The GWs are connected to the NS operator, located in Rome at the operator data center premises, where the database is also stored.

The database contains several indexes, each one representing a particular flow of information gathered from the network. For example, there are flows representing uplink and downlink packets, information exchanged between GW and the NS, or between EDs and NS, or packets which have been de-duplicated because they have been received several times. The latter case happens when different GWs insist in the same geographic area and packets reach the NS from these different GWs. All the studies have been developed over the “pre-deduplication” index; in such a way we are sure that all traffic analyzed is coming from the ED to the NS, passing through different GWs. The most relevant database fields, and the relative description, are reported in Table 1. A subset

of these fields will be extracted and pre-processed and they will represent the features for the profiling analysis performed in this work: this concept will be stressed in Sect. 5.

Table 1. The most important fields present in a pre-deduplicated device packet.

PARAMETER	DESCRIPTION
CHANNEL	Channel used to send the packet
CODR	Coding rate packet
CREATED_AT	Timestamp indicating the time when the entry has been created inside the database
DATR	SF and bandwidth of the packet
DEV_ADDR	Unique identifier of the device inside the network
DEV_EUI	Unique identifier of the physical device (None if the device is unknown)
FREQUENCY	Frequency of the sent packet in MHz
GATEWAY	GW MAC ADDRESS which received the packet
LSNR	Received signal to noise ratio of the packet
FCNT	Frame counter: counter increasing for each packet sent from ED; used evaluate error rate
RSSI	Received packet signal strength
SIZE	Packet size (bytes)
CRC STATUS	result: if 1 = passed if -1 = failed
TMST	Internal clock timestamp from gateway: used for synchronizing the downlink with the end transmission of the uplink to communicate response to end-device
UID	Unique identifier of the entry inside the database

4 k -means Algorithm and Best k Selection

As introduced in Sect. 1, the core profiling framework relies on k -means [15, 16], possibly the most famous data clustering algorithm. k -means is a partitional clustering algorithm [10, 18] and, as such, given a dataset $\mathcal{S} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ of n observations, it partitions the data into k non-overlapping clusters, i.e. $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_k\}$ such that $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset$ if $i \neq j$ and $\cup_{i=1}^k \mathcal{S}_i = \mathcal{S}$. k -means finds a (sub-)optimal partition of the data in such a way that the intra-cluster variance (also known as Within-Clusters Sum of Squares – WCSS) is minimized:

$$WCSS = \sum_{i=1}^k \sum_{\mathbf{x} \in \mathcal{S}_i} \|\mathbf{x} - \mathbf{c}^{(i)}\|^2 \quad (2)$$

where $\mathbf{c}^{(i)}$ is the centroid for cluster i , defined as the center of mass of the cluster itself. The k -means workflow can be summarized as follows:

1. select a set of k initial centroids (e.g., uniformly at random amongst the available data or by heuristics such as k -means++ [2])
2. assignment step: assign each data point to the closest centroid
3. update step: re-evaluate centroids for all clusters
4. loop 2–3 until convergence (e.g., centroids stop changing or a maximum number of iterations is reached).

Despite its simplicity, the number of clusters k to be returned is a parameter that must be personalized by the end-user and finding a suitable value is strictly problem- and data-dependent and hardly known a-priori. Typically, one tries several k candidates and selects the best value by studying the objective function in Eq. (2) and/or by means of internal validation indices [20]. Common strategies include:

The Elbow Plot [30] consists in plotting the WCSS as function of k and choose the first k value corresponding to the point where the curve become flat. The rationale behind this criterion is that is pointless to add more clusters if they do not give a better modelling of the data (the curve flattens since the WCSS does not change significantly)

The Davies-Bouldin Index [6] measures the intra-cluster separation against the inter-cluster variance. Let S_i be the statistical dispersion of cluster i , namely the average pattern-to-centroid distance, and let $M_{i,j}$ be the distance between centroids belonging to clusters i and j . For a clustering solution to be good S_i should be small (compact cluster), whereas $M_{i,j}$ should be large (different clusters are well far apart), hence for each pair of clusters one can define the following penalty score

$$R_{i,j} = \frac{S_i + S_j}{M_{i,j}} \quad (3)$$

and the Davies-Bouldin score for cluster i is defined as

$$DBI_i = \max_{j \neq i} R_{i,j} \quad (4)$$

Finally, the Davies-Bouldin score for the overall clustering solution is taken by averaging each cluster's score:

$$DBI = \frac{1}{k} \sum_{i=1}^k DBI_i \quad (5)$$

The Davies-Bouldin index is not bounded within a specific range; however, the closer to 0, the better.

The Silhouette Score [26] quantifies how-well each pattern has been assigned to its own cluster. For each point \mathbf{x}_i it is possible to define $a(\mathbf{x}_i)$ as the average distance between \mathbf{x}_i and all patterns in its own cluster. Similarly, it is possible to define $b(\mathbf{x}_i)$ as the nearest-cluster average distance of which \mathbf{x}_i is not a member. The silhouette score $s(\mathbf{x}_i)$ is defined as

$$s(\mathbf{x}_i) = \frac{b(\mathbf{x}_i) - a(\mathbf{x}_i)}{\max\{b(\mathbf{x}_i), a(\mathbf{x}_i)\}} \quad (6)$$

Finally, the silhouette score for the overall clustering solution is taken by averaging each point’s score:

$$s = \frac{1}{n} \sum_{i=1}^n s(\mathbf{x}_i) \quad (7)$$

Conversely to the Davies-Bouldin index, the silhouette score is by definition bounded in range $[-1, +1]$: the closer to $+1$, the better.

5 ED Profiling Through Data Clustering

5.1 Dataset Description

The main idea of this work is the study of the radio and network behaviour of the ED, and the goal is to cluster (group) EDs according to their behaviour. The dataset used in our analysis regards 3 months of activity over the LoRaWAN network (September–November 2018). The initial dump from the database counts more than 1 million packets and the following data filtering/cleaning operations have been performed:

1. all packets having NULL device identifier have been discarded: these packets correspond to device(s) which are either not managed by the operator or device(s) with unknown manufacturer;
2. all packets captured by a unique specific GW have been discarded: preliminary analyses showed that such gateway was managing a huge number of noisy packets with no valuable information.

This data filtering step dropped the number of available packets to 286,753 with 765 devices involved. In order to properly feed the k -means algorithm, the following pre-processing steps have been performed:

1. packets have been grouped by device;
2. each device has been mapped into an 11-length real-valued vector containing the following statistics amongst its packets (Table 1): the error rate (elaborated via the FCNT field), the average values of: SNR; RSSI; packet size; and inter-arrival time (i.e., number of hours between two packets), the mode values of: channel frequency; and SF, the number of packets sent during the night (12am–06am), during the morning (06am–12pm), during the afternoon (12pm–6pm) and during the evening (6pm–12am).

Amongst the available fields inside a LoRa device packet, we chose the aforementioned ones because they depict the radio aspect of the EDs: indeed, as introduced at the beginning of this Section and in Sect. 1, we recall that the target of our analysis is to profile EDs according to their radio and network behaviour. Thanks to this pre-processing stage, the considered devices have been cast into a 765×11 real-valued matrix, suitable for being processed by the k -means algorithm from Sect. 4.

5.2 Evaluation Results

In Sect. 4 it has been discussed that determining a-priori a suitable value for k (i.e., the number of clusters) is hard in many real-world applications. To this end, we considered several candidates $k = \{2, 3, \dots, 20\}$ and Fig. 4 shows the Davies-Bouldin Index, the Silhouette Index and the WCSS as function of k . By jointly considering the three indices, a suitable value of $k^* = 7$ has been chosen; indeed, the Silhouette is rather high (>0.9), the Davies-Bouldin index reaches its minimum value (<0.2) and $k = 7$ lies pretty much towards the end of the WCSS elbow, which can be easily seen for $k \in [3, 7]$.

For the chosen k^* , we further analyzed the clustering solution as returned by k -means. Indeed, the possibility of analyzing a-posteriori the contents of the returned clusters is one of the overwhelming advantages of cluster analysis, allowing a further knowledge discovery phase, possibly leaded by field-experts [8]. Specifically, for each cluster, we considered the behaviour of its ‘most central’ element (i.e., the element closest to the centroid) and Figs. 5, 6 and 7 show their most characteristic features.

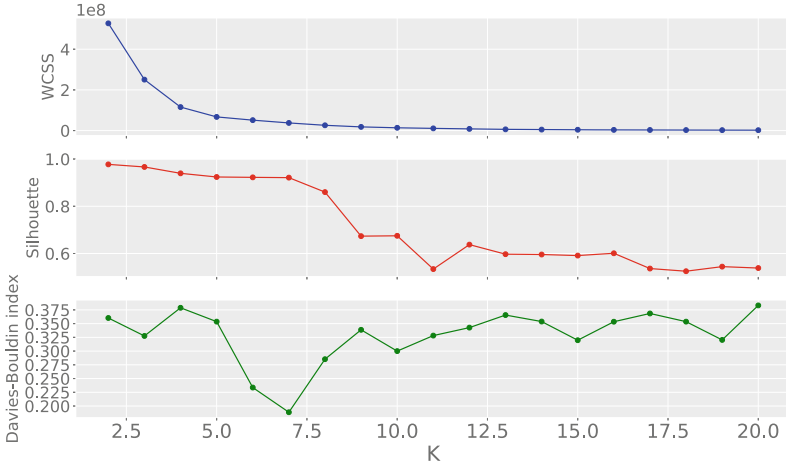


Fig. 4. WCSS, Silhouette and Davies-Bouldin indices as function of k .

Figure 5 shows the Probability Density Function (PDF) of the SNR for the ‘most central’ elements of the 7 different clusters. In the same way, Fig. 6 shows the PDF of the RSSI. Figure 7 shows the bar plot for the error rate values (Fig. 7(a)) and the number of packets sent (Fig. 7(b)). In all figures, device IDs have been anonymized and replaced by the ID of their respective cluster (reported in the figures legend) and marked with a unique color which consistently spans across Figs. 5, 6 and 7. For the sake of readability, legends are not included in Fig. 7.

By looking at Fig. 5, there exist three clear-cut profiles approximately bounded by $SNR < -10$, $SNR \in [-10, 10]$ and $SNR > 10$. Cluster 1 emerges

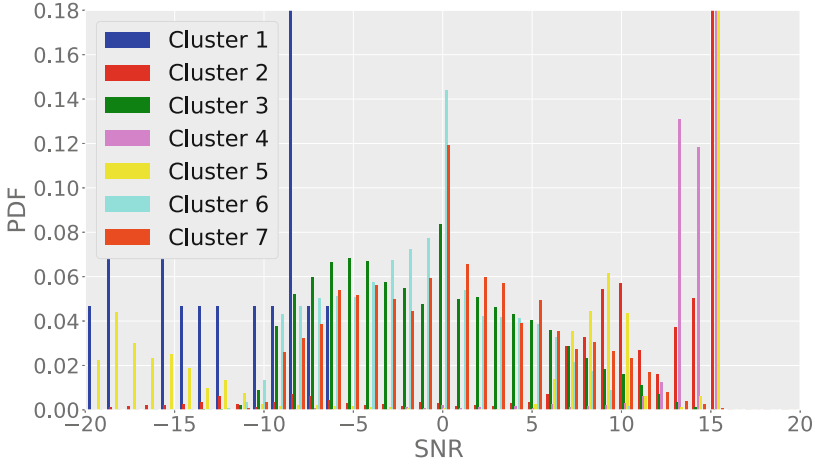


Fig. 5. PDF of the SNR for the ‘most central’ elements for the 7 clusters.

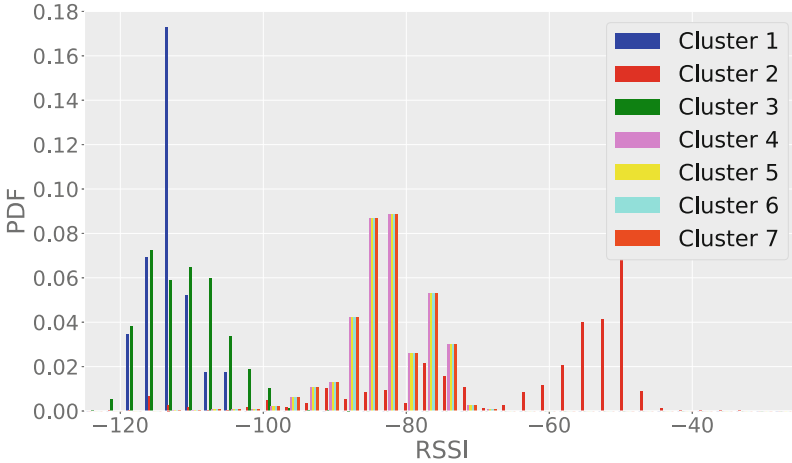


Fig. 6. PDF of the RSSI for the ‘most central’ elements for the 7 clusters.

to be in the first family, whereas the middle spectrum sees clusters 7, 6 and 3 as the most prominent ones and, finally, cluster 4 and 2 are identified by the highest SNRs. Interestingly, cluster 5 spans both the highest and the lowest SNRs. Clear-cut profiles also emerge in terms of RSSI (Fig. 6) bounded by $RSSI \approx -100$ and $RSSI \approx -70$: the leftmost part of the spectrum sees cluster 1 and 3 as the most prominent ones, whereas the rightmost part is dominated by cluster 2. All other clusters lie in the middle part of the spectrum. Figure 7 also shows completely different behaviours amongst the 7 clusters.

An interesting analysis consists in jointly considering these behaviours, which has been used as guideline to further characterize the radio properties of the

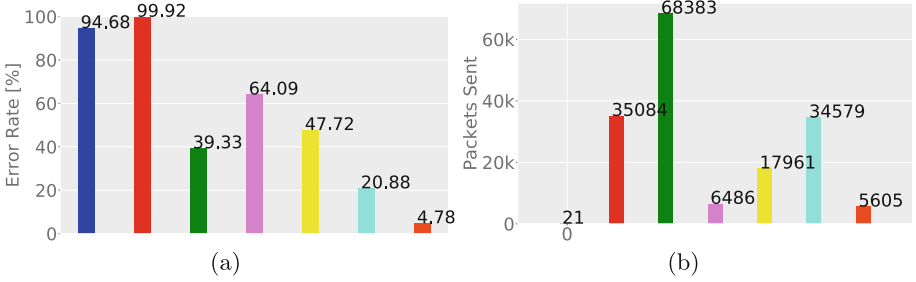


Fig. 7. Error rate (a) and number of packet sent (b) for the ‘most central’ elements for the 7 clusters

devices belonging to each group. Notably, field-experts at the network operator recognized some devices having precisely the characteristics presented in these histograms (e.g., the device with a large number of packets sent is, actually, a device which sends about three packets per day, so these clustering results are indeed in line with the network behaviour). At the same time, the cluster analysis was capable to find anomalies inside the network: for example, cluster 1 represents a set of devices which have high error rate (Fig. 7(a)) and small number of packets sent (Fig. 7(b)). This, along with low SNR values (Fig. 5), identifies a misbehaving of such devices.

6 Conclusion and Future Work

In this work, we have presented a study about the profiling of IoT devices. The study was performed on a real database that collects LoRaWAN packets received by a network deployed by an Italian operator. We tested our machine learning approach on 286,753 packets with 765 devices involved. In this work, we used the k -means algorithm in order to find suitable groups of devices sharing similar behaviour. Each device has been cast towards a suitable vector space by considering radio and network statistics amongst its packets. The soundness of the proposed clustering solution has been addressed by jointly considering three internal validation indices (WCSS, Silhouette and Davies-Bouldin), which also helped in tackling the problem of finding the best k value for the dataset at hand. Alongside the satisfactory results in terms of internal validation indices ($DBI < 0.2$ and $s > 0.9$), we further analyzed the resulting clusters by considering the distribution (in terms of either histograms or PDFs) of the most interesting features of the ‘most central’ elements of the clusters. With the help of field-experts, we were able to address the clustering solution in terms of knowledge discovery and the proposed approach has been demonstrated to be suitable also for anomaly detection purposes. Future research endeavours can consider the intrinsic structured nature of the data available within the LoRaWAN network for a more in-depth analysis. Indeed, in this work we considered basic statistics drawn from some features of the available packets, whereas one can perform

similar analyses on entire packets or sequence of packets by means of clustering algorithms such as k -medoids [19, 21] which do not necessarily require the input data to be in a vector form.

Acknowledgement. We thank UNIDATA S.p.A. who provided insight and expertise that greatly assisted our research, as well the access to a subset of the data for the analysis.

References

1. Aceto, G., Ciunzo, D., Montieri, A., Pescapé, A.: Mobile encrypted traffic classification using deep learning: experimental evaluation, lessons learned, and challenges. *IEEE Trans. Netw. Serv. Manag.* **16**(2), 445–458 (2019)
2. Arthur, D., Vassilvitskii, S.: K-means++: the advantages of careful seeding. In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007*, pp. 1027–1035. Society for Industrial and Applied Mathematics, Philadelphia (2007)
3. Barrachina-Muñoz, S., Adame, T., Bel, A., Bellalta, B.: Towards energy efficient LPWANs through learning-based multi-hop routing. In: *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, pp. 644–649 (2019)
4. Bhatt, P., Morais, A.: Hads: hybrid anomaly detection system for IoT environments. In: *2018 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC)*, pp. 191–196 (2018)
5. Croce, D., Gucciardo, M., Tinnirello, I., Garlisi, D., Mangione, S.: Impact of spreading factor imperfect orthogonality in LoRa communications. In: Piva, A., Tinnirello, I., Morosi, S. (eds.) *TIWDC 2017. CCIS*, vol. 766, pp. 165–179. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67639-5_13
6. Davies, D.L., Bouldin, D.W.: A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-1**(2), 224–227 (1979)
7. Del Campo, G., Gomez, I., Sierra, S.C., Martinez, R., Santamaria, A.: Power distribution monitoring using LoRa: coverage analysis in suburban areas. In: *Proceedings of the 2018 International Conference on Embedded Wireless Systems and Networks, EWSN 2018*, pp. 233–238 (2018)
8. Di Noia, A., Martini, A., Montanari, P., Rizzi, A.: Supervised machine learning techniques and genetic optimization for occupational diseases risk prediction. *Soft Comput.* (2019). <https://doi.org/10.1007/s00500-019-04200-2>
9. Hammerschmidt, C., Marchal, S., State, R., Pellegrino, G., Verwer, S.: Efficient learning of communication profiles from IP flow records. In: *2016 IEEE 41st Conference on Local Computer Networks (LCN)*, pp. 559–562 (2016)
10. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. *ACM Comput. Surv. (CSUR)* **31**(3), 264–323 (1999)
11. Kim, D.Y., Kim, S.: Data transmission using k-means clustering in low power wide area networks with mobile edge cloud. *Wirel. Pers. Commun.* **105**(2), 567–581 (2019)
12. Lueth, K.L., et al.: *State of the IoT & short-term outlook* (2018)
13. Kumar, A., Lim, T.J.: Edima: early detection of IoT malware network activity using machine learning techniques. In: *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, pp. 289–294 (2019)

14. Kurniabudi, K., Purnama, B., Sharipuddin, S., Stiawan, D., Darmawijoyo, D., Budiarto, R.: Preprocessing and framework for unsupervised anomaly detection in IoT: work on progress. In: 2018 International Conference on Electrical Engineering and Computer Science (ICECOS), pp. 345–350 (2018)
15. Lloyd, S.: Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**(2), 129–137 (1982)
16. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Oakland, CA, USA, vol. 1, pp. 281–297 (1967)
17. Marchette, D.: A statistical method for profiling network traffic. In: Proceedings of the Workshop on Intrusion Detection and Network Monitoring, pp. 119–128. USENIX Association, Berkeley (1999)
18. Martino, A., Giuliani, A., Rizzi, A.: Granular computing techniques for bioinformatics pattern recognition problems in non-metric spaces. In: Pedrycz, W., Chen, S.M. (eds.) *Computational Intelligence for Pattern Recognition*. SCI, vol. 777, pp. 53–81. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-89629-8_3
19. Martino, A., Rizzi, A., Frattale Mascioli, F.M.: Efficient approaches for solving the large-scale k -medoids problem. In: Proceedings of the 9th International Joint Conference on Computational Intelligence - Volume 1: IJCCI, pp. 338–347. INSTICC, SciTePress (2017)
20. Martino, A., Rizzi, A., Frattale Mascioli, F.M.: Distance matrix pre-caching and distributed computation of internal validation indices in k -medoids clustering. In: 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2018)
21. Martino, A., Rizzi, A., Frattale Mascioli, F.M.: Efficient approaches for solving the large-scale k -medoids problem: towards structured data. In: Sabourin, C., Merelo, J.J., Madani, K., Warwick, K. (eds.) *IJCCI 2017*. SCI, vol. 829, pp. 199–219. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-16469-0_11
22. Mostafa, B.: Monitoring internet of things networks. In: 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), pp. 295–298 (2019)
23. Muntean, V.H., Muntean, G.: A novel adaptive multimedia delivery algorithm for increasing user quality of experience during wireless and mobile e-learning. In: 2009 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, pp. 1–6 (2009)
24. Nivaashini, M., Thangaraj, P.: A framework of novel feature set extraction based intrusion detection system for internet of things using hybrid machine learning algorithms. In: 2018 International Conference on Computing, Power and Communication Technologies (GUCon), pp. 44–49 (2018)
25. Nömm, S., Başı, H.: Unsupervised anomaly based botnet detection in IoT networks. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 1048–1053 (2018)
26. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987)
27. Semtech: LoRa. EP2763321 from 2013 and U.S. Patent 7,791,415 from 2008
28. Sornin, N., Yegin, A., et al.: LoRaWAN 1.1 Specification (2017). <https://loralliance.org/resource-hub/lorawantm-specification-v11>
29. Tao, M., Ming, Y.C., Juan, C.: Profiling and identifying users’ activities with network traffic analysis. In: 2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS), pp. 503–506 (2015)
30. Thorndike, R.L.: Who belongs in the family? *Psychometrika* **18**(4), 267–276 (1953)

31. Verzegnassi, E.G.M., Tountas, K., Pados, D.A., Cuomo, F.: Data conformity evaluation: a novel approach for IoT security. In: 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), pp. 842–846 (2019)
32. Zhang, H.W., Sun, L., Zhang, H.: Research on data packets clustering algorithm in the wireless multiple hop network. *Appl. Mech. Mater.* **651**, 1905–1908 (2014)
33. Zhang, J., Chen, J.: An adaptive clustering algorithm for dynamic heterogeneous wireless sensor networks. *Wirel. Netw.* **25**(1), 455–470 (2019)



Experiences from Using Gamification and IoT-Based Educational Tools in High Schools Towards Energy Savings

Federica Paganelli²(✉), Georgios Mylonas³, Giovanni Cuffaro¹,
and Ilaria Nesi⁴

¹ CNIT, Firenze, Italy

giovanni.cuffaro@cnit.it

² Computer Science Department, University of Pisa, Pisa, Italy

federica.paganelli@unipi.it

³ Computer Technology Institute and Press “Diophantus”, Patras, Greece

mylonasg@cti.gr

⁴ Gramsci Keynes High school, Prato, Italy

ilanesi@alice.it

Abstract. Raising awareness among young people, and especially students, on the relevance of behavior change for achieving energy savings is increasingly being considered as a key enabler towards long-term and cost-effective energy efficiency policies. However, the way to successfully apply educational interventions focused on such targets inside schools is still an open question. In this paper, we present our approach for enabling IoT-based energy savings and sustainability awareness lectures and promoting data-driven energy-saving behaviors focused on a high school audience. We present our experiences toward the successful application of sets of educational tools and software over a real-world Internet of Things (IoT) deployment. We discuss the use of gamification and competition as a very effective end-user engagement mechanism for school audiences. We also present the design of an IoT-based hands-on lab activity, integrated within a high school computer science curriculum utilizing IoT devices and data produced inside the school building, along with the Node-RED platform. We describe the tools used, the organization of the educational activities and related goals. We report on the experience carried out in both directions in a high school in Italy and conclude by discussing the results in terms of achieved energy savings within an observation period.

Keywords: Internet of Things · Energy awareness · STEM education · Sustainability · Evaluation · Gamification

This work has been supported by the “Green Awareness In Action” (GAIA) research project, funded by the European Commission and the EASME under H2020 and contract number 696029. This document reflects only the authors’ views and the EC and EASME are not responsible for any use that may be made of the information it contains.

© Springer Nature Switzerland AG 2019

I. Chatzigiannakis et al. (Eds.): AmI 2019, LNCS 11912, pp. 75–91, 2019.

https://doi.org/10.1007/978-3-030-34255-5_6

1 Introduction

The Internet of Things (IoT) and smart cities have been two very active research fields during recent years, with a considerable amount of resources invested into building related infrastructures, creating large-scale smart city and IoT installations around the world. The capability of easily integrating cheap ubiquitous sensors into information systems has accelerated the growth of Ambient Intelligence (AmI) and the capacity of responding to specific situations according to context-awareness principia [20]. However, the question remains: how can we utilize such smart city and IoT deployments, in order to produce reliable, economically sustainable and socially fair solutions to create public value? This is especially true in the case of the educational domain, where it is more complex to integrate such solutions, given the restrictions in time and resources in school environments available for carrying out novel activities.

At the same time, there is an increasing interest in getting schools involved in raising awareness about climate change and energy efficiency. In this context, the importance of the educational community is evident, both in terms of size and future significance. Today's students are the citizens of tomorrow, and they should have the scientific and technological skills to respond to challenges like the climate change. The relevance of sustainable energy and energy saving behavior is gaining increasing interest in schools and in educational programs. Indeed, in the last decade schools have been the target of studies, education initiatives as well as energy efficiency actions in several countries. There has also been interest recently regarding the design of educational activities for energy awareness centered around IoT-enabled experimentation approaches. It is a means through which Europe can meet its goals, by equipping citizens, enterprise and industry in Europe with the skills and competences needed to provide sustainable and competitive solutions to the arising challenges [5].

In terms of research questions, which we were interested to answer through our work, the first one would be “how to engage and motivate end-user groups of students to participate in energy-saving educational activities”. The second one would be “whether IoT-based and data-driven educational interventions towards sustainability awareness actually work inside the classroom”. In other words, we wanted to look into the issue of motivating a school community using either more “soft” methods like gamification, or more technical hands-on ones, like IoT-based educational lab activities to educate students on the subject of sustainability.

Having in mind these questions, we discuss here our findings from applying gamification and competition mechanics in an Italian high school, with the purpose of getting the students more engaged into energy-saving activities in the context of a research project, entitled *Green Awareness in Action* (GAIA). With respect to related work, several gamification approaches exist for increasing awareness in energy sustainability topics [7]. In this context, the original contribution of our work consists in proposing a gamification and IoT-based approach designed to be integrated into the education cycle (e.g., introducing the concept of *Activity Class*) with a data-driven educational methodology leveraging either a fixed sensor infrastructure or sensors offered by cheap sensory boards rather

than designed as a standalone activity. On the other side, several experiments have been done with IoT-based activities targeting STEM education (e.g. [12]) without being linked to challenges like e.g., sustainability.

In this paper, we present the tools developed to support such approach, including an educational toolkit, comprising IoT devices and measurements data directly produced from school buildings. This toolkit allowed us defining educational activities within the school's computer science curriculum, leveraging IoT devices for environmental data acquisition and the integration with sensor measurements. It is worth noticing that our approach consists in leveraging an AmI approach to raise students' awareness on sustainability topics and motivate their behavior change instead of deploying building automation tools. We describe the organization of the activity and related curricula and educational goals to raise students' awareness in behaviour-based sustainability. Finally, we report on the experience carried out in both directions in a high school in Italy and conclude by discussing the results in terms of achieved energy savings in the observation period. The activities described here were conducted in this school during 2 school years, 2017-18 and 2018-19. As mentioned above, they include the application of both "soft" (gamification and competition) and "hands-on" (lab kit) mechanisms in order to engage high school students.

The paper is structured as follows. Section 2 discusses related work. In Sect. 3, we briefly describe the goals of the GAIA project and the real-world IoT infrastructure deployed in several school buildings and the list of tools developed within GAIA for supporting educational activities targeting behaviour-based energy savings. Section 4 discusses aspects related to gamification and competition, mostly related to activities in the schools during school year 2017-18, while Sect. 5 discusses aspects related to hands-on lab activities conducted during school year 2018-19. Our results indicate that both approaches can lead to interesting results; in our case, the increased engagement of the students led to both actual energy savings, as discussed in Sect. 6, and positive learning outcomes. Section 7 concludes the paper with insights into future work.

2 Related Work

The European Union is placing a strong focus on energy efficiency with initiatives like Build Up [8], a portal for energy efficiency in buildings. Overall, the percentage of school buildings among non-residential ones in Europe is around 17% [4]. Regarding the current state of the art in inclusion of sustainability and other related aspects in the educational domain, there is a lot of activity taking place with respect to inclusion of makerspace elements in school curricula, aided by the availability of IoT hardware as well. The work in [18] summarizes recent activity within the Maker Movement approach, presenting relevant recent findings and open issues in related research. Eriksson et al. [10] discuss a study stemming from a large-scale national testbed in Sweden in schools related to the maker movement, along with the inclusion of maker elements into the school curriculum of Sweden. Furthermore, there is a growing number of research projects

and activities that focus specifically on energy efficiency within the educational domain such as ZEMedS [1] and School of the Future [21]. Other recent projects like Entropy [2] target diverse end-user communities and do not focus specifically on the educational community. Moreover, several recent works focus on university curricula for teaching Internet of Things (IoT) leveraging a learning-by-doing and hands-on approach [6, 9, 13, 23], while the design of IoT-enabled educational scenarios in high and junior schools is less investigated. Porter et al. [19] argue that the lack of students' engineering experiences in primary and secondary education is in part due to the fact that very few teachers have an engineering/technology background and that the collaboration with universities and professionals would help coping with this issue. Along this direction, Gianni et al. [11] report on the usage of a toolkit [15] for rapid IoT application prototyping with a group of high school students. Analogously, [14] proposed the application in schools of a plug-and-play toolkit together with some suggestions for successful implementation of activities in high schools. An educational framework leveraging ubiquitous, mobile and Internet of Things technology for science learning in high schools has been proposed within the UMI-Sci-Ed project, while also investigating students' stance on IoT-enabled education activities [12].

There are examples of IoT-driven educational activities performed with the additional objective of increasing students' awareness of societal challenges. Tziortzioti et al. [22] designed and experimented data-driven educational scenarios for secondary schools to raise students' awareness of water pollution. However, these activities are typically conceived with an approach that cannot be easily applied and integrated into high school curricula. Mylonas et al. [17] proposed an educational lab kit and a set of educational scenarios primarily targeting primary schools for increasing energy awareness within the GAIA Project. With respect to gamification utilizing IoT in the context of sustainability, and specifically for energy and water, a recent survey on the subject is provided in [7]. However, although there are several examples of using gamification in this context, there has been little focus so far on the benefits of such an approach inside classrooms, an aspect we discuss in this work.

3 Overview of the Deployment Environment

The work presented here was conducted in the context of GAIA [16], a research project focusing on energy efficiency in educational buildings, employing behavioural change strategies, i.e., not using invasive techniques or retrofitting the buildings with actuators. This project produced a real-world multi-site IoT infrastructure comprising several school buildings in Greece, Italy and Sweden. The schools cover a range of local climatic conditions and educational levels (i.e., from primary to high school), as well as cultural settings.

Within this infrastructure, hereafter referred to as GAIA IoT Platform, a large number of IoT monitoring endpoints have been installed inside classrooms using heterogeneous hardware and software technologies, including different commercial hardware/sensor vendors, as well as open-source solutions. At each site,

the following types of measurements are periodically acquired: the power consumption of the whole building and selected rooms/areas, the environmental parameters of selected classrooms and/or laboratories (typically 5-minutes average values), and weather conditions and air pollution levels. These measurements, aggregated at different time granularity (e.g., 5-minute, hour, day, etc.), can be accessed through a set of REST APIs. A set of software and hardware artifacts have been produced within the project with the aim of experimenting different ways for raising students awareness on sustainability and energy consumption topics, also leveraging data and services provided by the GAIA IoT platform. In this work we focus on two tools and the activities designed around them:

- a web application (from now on called the Challenge) that serves as a playful introduction to sustainability and energy-related concepts for students. It uses gamification mechanisms to increase end-user engagement;
- an IoT-based educational lab kit that uses open-source technologies. It includes assembled devices and commercial IoT sensors and actuators to allow students to complete classes/tutorials regarding energy and sustainability.

In this work, we chose to focus on a 2-year experience carried out in a specific high school located in Italy, where both types of activities have been performed.

4 Gamification and Competition

In this section, we report on our experience in supporting the use of the GAIA Challenge in three classes of the target school. Overall, the Challenge is an online application aimed at students, designed to raise energy awareness and act as a playful introduction to sustainability aspects by leveraging gamification mechanics. The core of the application is a set of online Quests, grouped into five subject areas related to energy consumption reduction. The Quests are offered to students as steps of a “journey”, on top of a game “board”. This journey can also be enhanced with Class Activities, which are designed by teachers. For instance, Class Activities can consist in deeper investigation of some topics, energy-saving actions in the real-life, observation of monitoring data provided by the IoT platform, etc. The educational activities supported by the Challenge aim at motivating participants to engage in energy saving topics, by seeing their impact on the school facilities energy consumption over the course of the challenge and, finally, competing against other classes and schools.

Before introducing the Challenge to the classes, we performed a set of preparatory activities, such as disseminating advertising material in school areas and classrooms, training teachers through workshops, engaging the school principal and the technical staff. We also prepared brief eye-catching material, as an example of class activity created specifically for that school. Then, three classes of the high school started to participate to the activities. They conducted the Challenge Quests individually, while Class Activities were performed in groups.



Fig. 1. The “world” of the GAIA Challenge on the left, and the part where students see the schools’ score, trophies won and power consumption data on the right.

The students played the Challenge after a short introduction about the topics. Class Activities carried out by the classes consisted in analyzing monitoring data regarding the energy consumption and environmental comfort in the school facility, spotting possible rooms for improvement, devising and realizing ways of raising awareness in their school and family communities (e.g., news in the local newspaper, posts on the school web site, production of videos and presentations).

The students were divided into 5 groups and, to increase engagement, students underwent an evaluation. The teacher decided to assign a mark for the activity, accounted in the students’ final grade for the subject (physics in this case). The way this was computed was based on peer evaluation, and it was also proposed by the teachers of the school: first, a grade is assigned to the group as the average of the grades given by the other groups; then a grade is assigned for each member of a group by the other members of the same group. Student’s grade is the weighted average of these two evaluations (25% personal grade and 75% group grade). Such an evaluation mechanism boosted the individual contribution to the group activity and fair cooperation among group members.

All students of the three classes took part to the Challenge’s competition and they started to climb on the Challenge’s ranking of participating classes from all schools participating (over 20 schools in total). They also produced some snapshots like animated GIFs, videos and presentations. Overall, the level of engagement due to competition achieved in the school was very high. This was a successful outcome, which, however, can be associated to some unexpected and risky effects, as we experienced in the final days of the competition. Indeed, the classes began to *continuously monitor* the ranking and at a certain point in time they began arguing on fairness in the Challenge score and behavior

of competitors. This issue was raised when they noticed that after months of activities, another team surpassed them in the overall score. Essentially, what the students then tried to do amounted to “reverse engineering” the way that the scores were calculated in the Challenge. They basically started to monitor what the other teams were doing, and whether actions from their side had any tangible effect on their school’s overall score. At some point, they minimized the set of possible cases and scenarios for the way scores are calculated. This turn of events could be summarized as an ideal for the Challenge: our end-users were more than just engaged, they were *thrilled* to participate and out-compete other schools, which they hadn’t even heard of before.

Another aspect was that they assumed the way the other school surpassed them in the score was a “trick”, suggesting to us that the score counting method was “unethical”. They noticed that 3 new users were added to a competing school, thus contributing extra points to the overall class score. They were very sensitive to this issue and stated that they were also ready to cancel all the project-related activities, since they perceived the process as anti-pedagogical, or even unethical. From their point of view, they were not complaining because they were no longer the best team, but against what they perceived as not fair play. The students also proposed a solution to the score “issue”, as perceived by them. Essentially, the problem was that there was an assumption from the students’ side that new students could be added to a certain school after an initial period and that all students from the other schools had registered early on. As a way to counter such complaints in the future, we setup some measures to reduce the probability of cheating behaviours, namely: (i) registrations are allowed only for a limited period, teachers may request an extension providing a motivation, (ii) the teacher’s guide to the Challenge has been enhanced asking the teacher to control the identity of users registered to their class.

In terms of other overall comments about the contribution of gamification to the engagement, in several of the other schools participating in the project we saw increased engagement. Moreover, during 2 school years (2017-18, 2018-19) we announced 2 “competitions”, where all schools in the project were called to participate and for certain categories they should use the Challenge to score more points or create content to share with their peers. As a general observation, during the time period that the competitions ran, there were very easily identifiable spikes in the end-user activity in the Challenge, as seen in Fig. 2.

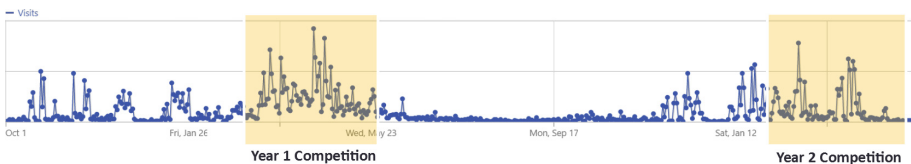


Fig. 2. A chart depiction of the end-user visits to the Challenge, with the time periods of 2 competitions marked and showcasing the effect on engagement.

5 IoT-Enabled Educational Activity

We focus now on the experience carried out in the design and experimentation of an “hands-on” IoT educational activity with the twofold objective of (i) increasing students’ awareness on the relevance of behavior change for achieving energy savings through a data-driven approach, thanks to real-world data gathered by the GAIA IoT platform, while also (ii) integrating such IoT-enabled experiment into a upper secondary school curriculum. The second aspect, i.e. designing the activity so that it could fit within the educational program of at least of one the subjects offered to students, was a key element to encourage teachers in devoting effort to contribute to the design and experimentation as well as in actively engaging students in proficiently taking part to the activity. The activity has been mainly designed to fit in the computer science curricula, although it can be further developed with actions carried out in the framework of additional subjects, such as sciences and physics. More specifically, the activity has been designed as a set of computer science lectures for a class of students of the 4th year of a Scientific Lyceum in Italy. The learning objectives for that year are defined at high level by the Ministry of Education¹ and then customized by each teacher. They generally consists in: learning a programming language, basics of coding, data modeling and tool for data access, manipulation and persistence, web programming. Moreover, the use of the acquired knowledge to support a study on topics in science and physics subjects is also welcomed.

In this context, the activity was designed to develop the theme of sustainability and energy awareness by first providing students with monitoring data gathered from the environment they live in (e.g., the school hall and the computer science laboratory) by using the sensor infrastructure deployed within the GAIA Project in a selected set of rooms and areas in the school (see Sect. 3) and a sensor board (called *LabKit*) that was given to students to gather environmental parameters in some of the remaining areas. Thanks to these tools, first, the students have access to a real-world dataset, which they can learn to manipulate and query through appropriate tools in order to derive meaningful data about the school environment they live in (energy consumption, environmental comfort, etc.). Based on such findings, students, with the help of the teachers, can decide on actions for further investigation (e.g., changing their behaviour to save energy, studying the factors that influence the comfort in school environments, etc.). Second, they can enhance the experience with programming tools to develop programs for data manipulation, sharing and visualization.

The material we prepared for computer science teachers consists in: the *LabKit* sensor board, documentation and examples for using a programming environment for developing programs for pushing sensor data from the *LabKit* to the GAIA IoT platform and simple web applications for data visualization. Hereafter, we describe the *LabKit* and the programming environment selected and extended for its usage within the GAIA Project and, then, report on the experience with a computer science class of 22 students.

¹ <https://www.miur.gov.it/liceo-scientifico-opzione-scienze-applicate> (in Italian).

5.1 Lab Kit and Programming Tools

The *LabKit* has been assembled utilizing the following components to minimize costs and ease replicability (shown in Fig. 3):

1. Raspberry Pi model 3B or 3B+: Raspberry Pi v3;
2. GrovePi, which is an add-on board that couples to the Raspberry to ease the connection of external sensors;
3. GrovePi Sensors, i.e., light, temperature, humidity, sound sensors, together with buzzer, LEDs and a Button; and
4. An LCD screen, allowing an immediate feedback on gathered measurements.



Fig. 3. Sample photos of the Raspberry Pi-based IoT hardware used in classrooms.

We chose Node-RED [3] as the environment to be used by students, since it leverages a flow-based programming environment for easily interconnecting hardware devices, online services and developing IoT applications. Indeed, Node-RED is a visual tool that allows users with minimal programming skills rapidly assembling and deploying an IoT application, and can be used in schools for performing simple experiments with sensors and IoT data processing in classes. Integrating Node-RED with the APIs of the GAIA IoT platform, and, optionally, local sensor kits, thus allows enriching the educational activities with the use of real sensor data, while leveraging a tool supported by a wide open source community and a rich documentation.

Node-RED comes with a set of ready-to-use customizable nodes allowing to design and deploy simple applications by simply dragging and dropping nodes from the palette and connecting them according to the desired flow. On the other side, Node-RED can also be easily extended by developing and adding new nodes to the repository. This extensibility provides a flexible support to the design of educational activities at a customizable difficulty level, which can be adapted to the class level and syllabus. For instance, teachers may assign tasks

to students for programming a custom data processing function or graphical widget and add them to Node-RED. In order to foster a data-driven approach, not merely limited to local usage of measurements gathered through the *LabKit*, we developed an additional Node (called *GaiaNode*) to access sensors and API of the GAIA IoT Platform as data sources in Node-RED. The software and documentation is available as open source on GitHub. The *GaiaNode* plug-in for Node-RED is a set of nodes that allows interacting with the GAIA IoT platform to retrieve measurements gathered by the fixed sensor infrastructure deployed in the schools involved in the GAIA Project, as well as pushing values of the *LabKit* sensors (e.g., Raspberry Pi sensors) into the platform.

5.2 Data-Driven Education for Energy Awareness

Hereafter, we describe one of the IoT-enabled educational activities that targeted high schools and integration with existing computer science curricula. The pedagogical goals aiming at increasing students awareness on energy topics are: awareness, observation, experimentation and action. The main steps of this IoT-enabled education activity are summarized in Table 1. The activity has been designed by researchers with a computer networks background, with the help of the computer science professor of the high school. A total of 22 high school students participated in the activity, carried out weekly in a 2-hour computer science class slot from February to end of April 2019.

Students chose to monitor the temperature of their computer science lab, since they experienced a too high and uncomfortable heat. The availability of the *LabKit* allowed them to monitor the conditions in the lab and correlate with outdoor conditions retrieved through the GAIA IoT Platform. They measured very high temperature values (in the range of 25 – 30°C) also in cold days and during night, when heating was supposed to be off. They also analyzed these data while varying the room conditions (windows on/off, curtains open/closed).

Since radiators in the laboratory were not equipped with thermostatic valves, they couldn't turn their observations into direct energy saving actions (e.g., regulating radiators). As an outcome of the discussion on Day 9 they elicited a set of questions and energy-saving proposals and decided to submit them to the school principal. This resulted in a 20-minutes discussion with the principal on pragmatic actions for guaranteeing comfort while achieving energy savings. The discussion was initially focused on the experimental findings in the computer science laboratory and, at the end, was extended to other critical areas of the school. The discussion ended up with a set of actions to be performed by the school principal and ideas for follow-up activities to be performed by students.

5.3 Evaluation

We gave a questionnaire to students to assess their satisfaction and engagement. The questionnaire is shown in Table 2 and was derived taking into account similar surveys in related work (e.g., [11, 12]). Answers were given on a Likert scale (1–5). It was submitted to students after the end of the activity. Figure 4 shows the

results. Responses were mostly positive about the satisfaction and engagement in the activity (i.e., Q1,Q2,Q7,Q9 and Q10). However, responses related to easiness of the activity (Q6, Q8) suggest the need for improvements (e.g., distributing activities across a longer span, additional documentation/tutorials).

Table 1. Template of IoT-enabled educational activity in high schools

Day	Activity description	Educational goal
1	Introductory seminar about energy consumption awareness and carbon footprint calculation	Introduction
2	Introduction to the IoT. Definition and examples of deployment and applications from the Web and the GAIA Project	Understanding the concept of IoT and related impact in the everyday life
3	Introduction to Node-RED and Flow-based programming. Notion of node, flow and deployment in Node-RED	Position Flow-Based Programming using Node-RED for programming simple applications for data manipulation
4	Design of basic Node-RED flow examples. Use of the <i>GaiaNode</i> plugin for accessing IoT measurements. Retrieving energy consumption data of the school	Extending Node-RED with new nodes. Observation of measurements of energy consumption and environmental parameters in the school
5	Configuration of Raspberry and temperature sensor to monitor the temperature in a selected area (e.g., the computer science laboratory)	Learning setting up and configuring sensor and computing hardware devices. Analyzing and processing measurement data through a spreadsheet
6	Develop a Node-RED flow application for creating a virtual sensor resource and pushing <i>LabKit</i> measurements into the GAIA Project platform	Learning programming a virtual sensor application pushing sensor measurements into the GAIA Project IoT platform
7	Develop a web-based Dashboard application to visualize temperature values	Developing programs for data access and visualization leveraging web protocols
8	Analysis of monitoring data in different conditions of the lab (lights on/off, windows open/closed, etc.	Experimentation: taking some actions and analyzing the impact on the environment
9	Discussion on findings and plan of short-term/long-term actions for further investigation, development, energy saving etc.	Analyzing data, finding issues and countermeasures, assigning priorities
10	Action (objective and duration of actions to be decided by students)	It depends on the action decided by the class

Table 2. Students' Questionnaire

ID	Question
Q1	I am satisfied with the activity
Q2	I am pleased with the activity
Q3	The activity was easy
Q4	The process of the activity was clear and understandable
Q5	I was able to follow the tasks of the activity
Q6	I have the knowledge and ability to follow the tasks of the activity
Q7	Attending the activity was enjoyable
Q8	Attending the activity was exciting
Q9	I was feeling good in the activity
Q10	I found the activity useful
Q11	The activity improved my capabilities in science and technologies
Q12	I liked to observe and use the data and measurements
Q13	I liked the lab activity with Node-RED and the Raspberry Pi
Q14	I learned something new by observing and using the data and measurements
Q15	I learned something new in the lab activity with Node-RED and the Raspberry Pi

Finally, since this activity was part of the computer science education program, at the end of the activity the students took a written exam for assessment that comprised 5 free-text questions (two questions on IoT and flow-based programming concepts, two on Node-RED usage for programming applications, and one on solutions for sustainability and energy awareness in the school). Students' scores were: 2 excellent, 3 good, 6 satisfactory, 7 sufficient, 4 insufficient. On average, the class performed better than previous class exams. According to the teacher, this was probably due to the fact that previous exams aimed to verify knowledge acquired through traditional learning methodology on a larger body of content, while our IoT activity was more engaging and part of the content was produced by the students themselves. In addition, the activity was done in groups and collaboration was an incentive for students to perform better and meet deadlines, as to not damage their classmates. As a final note, the teacher also said that the activity was successful in consolidating relations among students and between class and the teacher. The students demonstrated increased engagement compared to lectures in previous months, thanks also to the awareness that their work would have impact on their school environment. Moreover, the teacher suggested that the timeplan was too strict and the activity would benefit from additional time (e.g., 3 h a week instead of two) devoted to the activity. In that specific case, this would mean performing part of the activity

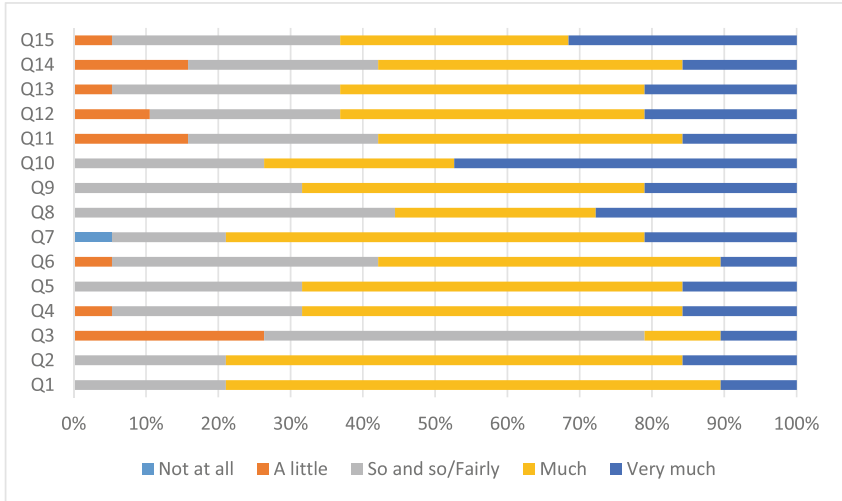


Fig. 4. Results on the questionnaire related to the learning outcome and the experience

within other subjects (physics or science), thus requiring the cooperation of a group of teachers and the enhancement of interdisciplinary aspects.

6 Behaviour-Based Energy Savings

Since the common objective of the educational activities presented above is energy awareness, here we report on an activity performed by students targeting energy savings in a school environment through behaviour change. The following is an example of possible decisions that students can take as an effect of awareness improvement through learning, observation and experimentation steps.

Leveraging measurements for their school made available by the GAIA IoT platform, the class decided to focus on the lighting of the main hall as the use-case for energy savings. With respect to luminosity, there is a minimum recommended value of 150 lx for areas like indoor halls. Luminosity sensors are installed, but given that they produce measurements that are highly related to their orientation, which is not optimal, the students had to calibrate the values they saw through the system. Making a rough estimation, students set a threshold of 400 lx for the values produced by the sensors that they thought it corresponded to “good enough” lighting. Figure 5 displays the measurements for power consumed by lights in the hall and luminosity, with the addition of the 400 lx threshold (horizontal line marked in red). Also highlighted in the figure is the interval during which luminosity in the school hall is above the threshold.

It is evident that between 10:00AM and 5:00PM lights should be turned off. This was a recurring situation in this school for months, due to its location (Italy) and orientation; i.e., it is not something that is observed for a single day

or over a short time period. The next step was to act according to the plan for turning off the unnecessary lights, while also making sure not to leave any part of the hall in the dark. Lighting should be turned off for sufficient time, in order to be able to observe the change in the data. It was convenient to calculate the average values of the lighting system during a “normal” baseline period and after the intervention. The school analyzed the new data regarding power during the period in which the light was turned off. With the lighting configured as usual, power consumption is approx. 4.9 kW. When the school acted to keep active only what is necessary, the power consumption decreased to 1.9 kW, thus saving 3 kW in the process. This practically means that 21 kWh could be saved during a single day, considering the 7 h of the interval during which this issue was identified. With such data in hand, students performed simple actions for raising awareness in the school staff for switching off the lights in the hall when not necessary and involved their schoolmates in similar actions in classrooms and laboratories.

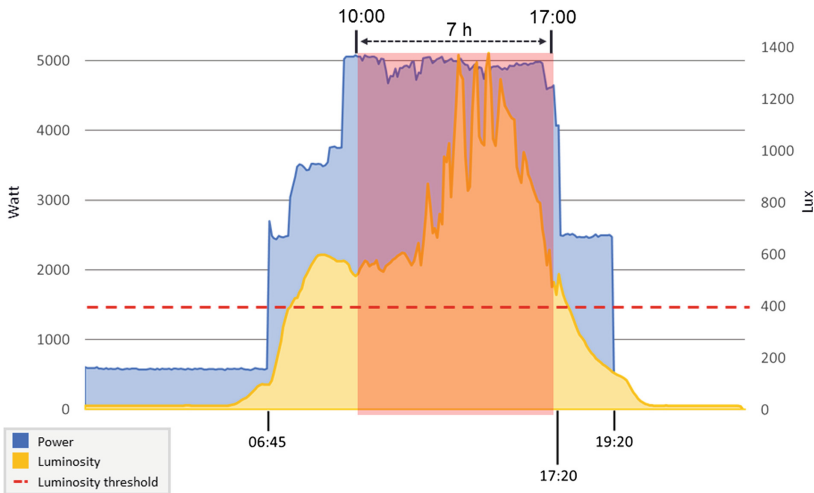


Fig. 5. An illustration of the period (highlighted) in which there is a waste of electricity. (Color figure online)

The potential energy savings analyzed in the previous steps pushed students to act. They created a set of signs (Fig. 6) to help school staff remember which switches can be turned off when natural light is enough. They also designed a poster placed in all rooms equipped with a projector: “Please shutdown the computer and the projector when not in use”. Also, one of the classes involved produced a short video to encourage friends and families to join the “battle for environmental care”. This video gives simple advices for saving energy and decreasing pollution. Based on achievements due to the short- and medium-term activities, the school wanted to support students in taking further measures

within a longer time span, to obtain the best results in terms of energy efficiency. Students periodically observed and analyzed the impact that these changes have in the long term and monitored progress toward achieving their objectives.

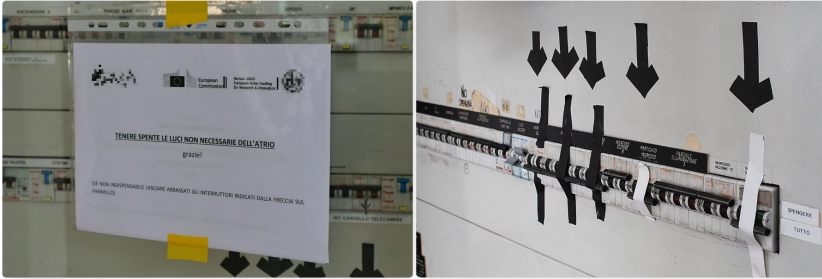


Fig. 6. Examples of simple interventions that the students made in the building.

7 Conclusions and Future Work

The educational community is one of the most interesting target groups for sustainability and energy-related activities. The successful introduction of such activities into the curricula of schools in Europe is still an open issue. We presented our experiences from utilizing mechanisms such as competitions, gamification and IoT-based, hands-on lab activities to increase the engagement of students at a high school in Italy. Our results, produced during 2 school years, provide some interesting insights regarding the creative ways the aforementioned mechanisms can be used to trigger the interest of high school students. The inclusion of competition and gamification aspects with students as end-users can greatly increase their engagement, especially when having groups/schools competing with each other. This helped us to distill some lessons:

- Direct and informal support to teachers: teachers were the gateway to our end-users; having gained the trust and attention of teachers is the first step to establishing a connection with the students as well.
- Provide short and captivating material: schools tend to have little time available to dedicate to extra-curricular activities, information should be as engaging and codified as much as possible.
- Co-design of tailored activities: teachers have to be involved in tailoring the proposed activities by taking into account their background and syllabus.
- Competitiveness is key for engagement but has to be handled with care: in our case, a rapid increase in engagement came close to backfire.

With respect to our future work, we plan to continue this line of research by applying our tools to school communities in other European countries.

References

1. ZEMedS project (Zero Energy MEDiterranean Schools). <http://www.zemeds.eu>
2. Entropy project. <http://entropy-project.eu>
3. Node-RED web site. <https://nodered.org/>
4. Europe's buildings under the microscope: A country-by-country review of the energy performance of buildings. Technical report, Buildings Performance Institute Europe (BPIE) (2011). ISBN: 9789491143014
5. Science europe roadmap (2013). <https://www.scienceeurope.org>
6. Abbasy, M.B., Quesada, E.V.: Predictable influence of IoT (Internet of Things) in the higher education. *Int. J. Inf. Educ. Technol.* **7**(12), 914–920 (2017)
7. Albertarelli, S., et al.: A survey on the design of gamified systems for energy and water sustainability. *Games* **9**(3), 38 (2018). <https://doi.org/10.3390/g9030038>
8. European Web Portal for energy efficiency in buildings. <http://www.buildup.eu>
9. Dobrilovic, D., Zeljko, S.: Design of open-source platform for introducing internet of things in university curricula. In: *IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pp. 273–276, May 2016
10. Eriksson, E., Heath, C., Ljungstrand, P., Parnes, P.: Makerspace in school - considerations from a large-scale national testbed. *Int. J. Child-Comput. Interact.* **16**, 9–15 (2018)
11. Gianni, F., Mora, S., Divitini, M.: Rapid prototyping internet of things applications for augmented objects: the tiles toolkit approach. In: Kameas, A., Stathis, K. (eds.) *AmI 2018*. LNCS, vol. 11249, pp. 204–220. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-03062-9_16
12. Glaroudis, D., Iossifides, A., Spyropoulou, N., Zaharakis, I.D.: Investigating secondary students' stance on iot driven educational activities. In: Kameas, A., Stathis, K. (eds.) *AmI 2018*. LNCS, vol. 11249, pp. 188–203. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-03062-9_15
13. He, J., Lo, D.C.T., Xie, Y., Lartigue, J.: Integrating internet of things (IoT) into stem undergraduate education: case study of a modern technology infused courseware for embedded system course. In: *2016 IEEE Frontiers in Education Conference (FIE)*, pp. 1–9, October 2016. <https://doi.org/10.1109/FIE.2016.7757458>
14. Katterfeldt, E.S., Cukurova, M., Spikol, D., Cuartielles, D.: Physical computing with plug-and-play toolkits: key recommendations for collaborative learning implementations. *Int. J. Child-Comput. Interact.* **17**, 72–82 (2018)
15. Mora, S., Gianni, F., Divitini, M.: Tiles: a card-based ideation toolkit for the internet of things. In: *Proceedings of the 2017 Conference on Designing Interactive Systems*, pp. 587–598. ACM (2017)
16. Mylonas, G., Amaxilatis, D., Chatzigiannakis, I., Anagnostopoulos, A., Paganelli, F.: Enabling sustainability and energy awareness in schools based on IoT and real-world data. *IEEE Pervasive Comput.* **17**(4), 53–63 (2018). <https://doi.org/10.1109/MPRV.2018.2873855>
17. Mylonas, G., Amaxilatis, D., Pocero, L., Markelis, I., Hofstaetter, J., Koulouris, P.: An educational IoT lab kit and tools for energy awareness in european schools. *Int. J. Child-Comput. Interact.* **20**, 43–53 (2019)
18. Papavaslopoulou, S., Giannakos, M.N., Jaccheri, L.: Empirical studies on the maker movement, a promising approach to learning: a literature review. *Entertainment Comput.* **18**, 57–78 (2017). <https://doi.org/10.1016/j.entcom.2016.09.002>
19. Porter, J.R., Morgan, J.A., Johnson, M.: Building automation and IoT as a platform for introducing STEM education in K-12. In: *ASEE Annual Conference and Exposition* (2017)

20. Ricciardi, S., Amazonas, J.R., Palmieri, F., Bermudez-Edo, M.: Ambient intelligence in the internet of things. *Mob. Inf. Syst.* **2017**, 1–3 (2017)
21. Towards Zero Emission with High Performance Indoor Environment. <http://school-of-the-future.eu>
22. Tziortzioti, C., Andreetti, G., Rodinò, L., Mavrommati, I., Vitaletti, A., Chatzigiannakis, I.: Raising awareness for water pollution based on game activities using internet of things. In: Kameas, A., Stathis, K. (eds.) *AmI 2018. LNCS*, vol. 11249, pp. 171–187. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-03062-9_14
23. Zhamanov, A., Sakhiyeva, Z., Suliyev, R., Kaldykulova, Z.: IoT smart campus review and implementation of IoT applications into education process of university. In: *13th International Conference on Electronics, Computer and Computation (ICECCO)* (2017)



IL4IoT: Incremental Learning for Internet-of-Things Devices

Yuanyuan Bao^(✉) and Wai Chen

China Mobile Research Institute, Beijing, China
baoyuanyuan@chinamobile.com, wai.w.chen@gmail.com

Abstract. Considering that Internet-of-Things (IoT) devices are often deployed in highly dynamic environments, mainly due to their continuous exposure to end-users' living environments, it is imperative that the devices can continually learn new concepts from data stream without catastrophic forgetting. Although simply replaying all the previous training samples can alleviate this catastrophic forgetting problem, it not only may pose privacy risks, but also may require huge computing and memory resources, which makes this solution infeasible for resource-constrained IoT devices. In this paper, we propose IL4IoT, a lightweight framework for incremental learning for IoT devices. The framework consists of two cooperative parts: a continually updated knowledge-base and a task-solving model. Through this framework, we can achieve incremental learning while alleviating the catastrophic forgetting issue, without sacrificing privacy-protection and computing-resource efficiency. Our experiments on MNIST dataset and SDA dataset demonstrate the effectiveness and efficiency of our approach.

Keywords: Incremental learning · Catastrophic forgetting · Internet of Things · Continuous learning · Autoencoder · Knowledge base

1 Introduction

With the development of information network, the popularity of Internet-of-Things (IoT) is an irreversible trend. There are very large numbers of devices being employed to support the vision of IoT. ABI Research estimates that there will be more than 30 billion devices connected to the Internet-of-Things by 2020. Just as Wu et al. [1] stated, only connected is not enough, IoT devices should have cognitive capability to learn, think and understand both physical and social worlds. In recent years, the cognitivity demands for IoT are becoming more and more urgent, and empowering the IoT with learning ability is the future direction.

Considering that the IoT devices are often deployed in highly dynamic and uncontrolled environments, it's impossible that the devices can access the training data of all classes at same time. That is, the IoT devices should be able to incrementally learn about classes, when training data for them becomes available

continuously. Although machine learning is an instrumental tool for the intelligence empowerment, which has been widely used in numerous vertical market segments, the current machine learning approach is probably not sufficient for learning tasks incrementally. As Liu et al. [2] stated, standard machine learning is isolated learning, which does not retain the knowledge learned from the previous tasks. And this is also called catastrophic forgetting, referring to the phenomenon when learning a sequence of tasks, the learning of each new task may cause the forgetting of the knowledge learned from the previous tasks. Anti-thetically, humans have the ability to incrementally learn over time by acquiring new knowledge while retaining previously learned knowledge. Such an incremental learning procedure has represented a challenging problem for building a truly intelligent IoT system.

In this paper, we attempt to design a lightweight incremental learning framework for resource-constrained IoT devices [3, 4]. Specifically, we focus on the case where a sequence of tasks arrives continually as shown in Fig. 1. In each training session, there is only one task arriving, with totally different classes within the task from the previous tasks. Then the aim is to incrementally learn new tasks and obtain a competitive model for all the tasks, when the corresponding training samples become available. For example, there is an IoT device deployed in user’s house, which can achieve activity recognition by adopting an already-trained model. It is assumed that the trained model can recognize activities, such as *sitting* and *standing*. When the user performs some new activities, such as *lying* and *jumping*, the IoT device is expected to incrementally learn to classify these newly-encountered activities without forgetting the old classes of activities. That is, we wish to refine the model not only to precisely recognize new classes, but also not to forget old classes. Although we can refine the recognition model by retraining the model offline based on combining the samples of new classes with the former training samples, this solution has several disadvantages: (1) the refinement will be seriously lagging because the retraining based on the total training samples may require substantial time (not near real-time); (2) the refinement may pose privacy risks due to that it requires the former training samples; (3) the retraining costs lots of computational and storage resources, which makes it not a scalable solution and unacceptable for the resource-constrained IoT devices.

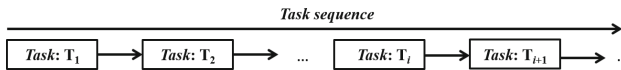


Fig. 1. An example of incremental learning for IoT devices

Several studies have proposed incremental learning methods to lessen the effect of catastrophic forgetting, e.g., incremental classifier and representation learning (iCaRL) [5], generative replay (GR) [6], learning without forgetting (LwF) [7], elastic weight consolidation (EWC) [8], gradient episodic memory

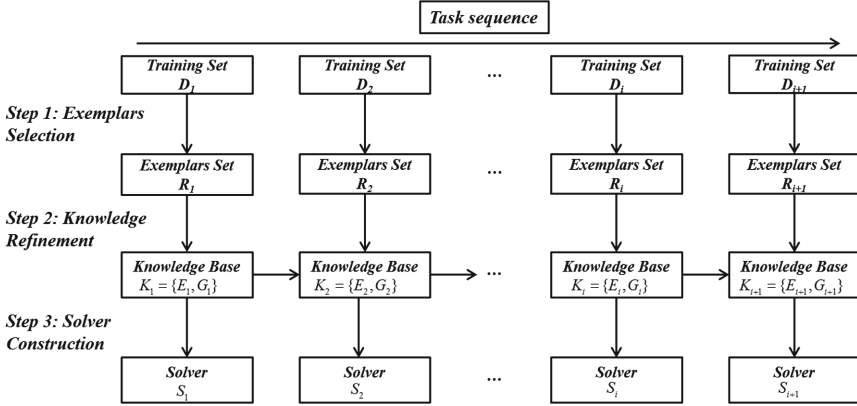


Fig. 2. IL4IoT framework. Two main components: knowledge base and solver. Three key steps: exemplar selection, knowledge refinement and solver construction.

(GEM) [9], etc. However, there are some disadvantages in the existing methods. For example, iCaRL chooses the exemplar sets and transfers to the consecutive task, which still causes the privacy risks. GR transfers the GAN network to assist the learning in consecutive tasks. However, because of the current limitations of GAN networks, purely GAN based approaches are not competitive yet as will be discussed in the following section.

In view of the disadvantages of the existing approaches and the characteristics of IoT devices [10], we propose a lightweight incremental learning framework as shown in Fig. 2. In order to retain the previously learned knowledge and assist the knowledge accumulation, in each training session, there are two cooperative parts: a knowledge-base and a task-solving model (called as solver). In our framework, knowledge-base consists of encoded exemplars and their decoder. With the task sequence, the knowledge-base will be refined continually. Then based on the continually refined knowledge-base in each stage, we can construct the solver for each stage. In order to reduce the computational complexity, we leverage autoencoder to map the original data into latent representation, based on which we conduct exemplar selection (step 1). In order to further reduce the computational complexity and protect the data privacy, we conduct knowledge refinement based on knowledge extraction, exemplars reduction and exemplars combination (step 2). Finally, based on the refined knowledge, we perform the solver construction (step 3).

The paper is organized as follows. In Sect. 2, we introduce related methods for incremental learning. In Sect. 3, we propose our incremental learning framework - IL4IoT and thoroughly describe the main mechanisms. In Sect. 4, we describe the experimental settings and evaluate our method using two benchmark datasets. In Sects. 5, we conclude this paper and discuss future work.

2 Related Works

Catastrophic forgetting [11] poses particular challenges for continual learning due to the fact that the knowledge learned from previous tasks will be lost when learning the new task. To deal with catastrophic forgetting (*CF*), many approaches have been proposed, which can be divided into two categories: requiring old data or not requiring old data.

The methods in the first category require part or all of the old data. McClosky and Cohen [11] explored the use of rehearsal in order to prevent forgetting. During rehearsal, training samples are randomly picked from old training sessions and then they are mixed with the ones in the new training session. Robins [12] proposed pseudo rehearsal, in which forgetting is prevented by generating random samples and then passing them through the network to obtain their labels. These examples along with their labels are then used during rehearsal. In order to retain the knowledge of previous classes, Rebuffi et al. [5] conduct prioritized exemplar selection, store exemplars for data representation updating and achieve the classification by a nearest-mean-of-exemplars rule. Lopez-Paz and Ranzato [9] propose Gradient Episodic Memory where they solve forgetting problem by storing data from previous training sessions and performing gradient update such that the error on previously learned tasks do not increase. The key difference between our method and these prior techniques is that we do not keep any real data of the old classes. In real world applications, due to the privacy and legal concerns and tight resource constraints of the IoT devices, real data may not be allowed to be stored for a long period of time. Since our method conducts knowledge extraction based on exemplars, the knowledge we retain is only the encoded exemplars and decoder network, which are not related to the data privacy. Therefore, our method not only can protect the data privacy, but also it can save the computational and memory storage costs of the devices.

The methods in the second category do not require any old data. Li and Hoiem [7] introduce a method named Learning without Forgetting (*LwF*) wherein by optimizing both the pseudo-training data of the old task and real data of the new task, LwF attempts to prevent catastrophic forgetting and is promising where the properties of the pseudo training set are similar to the ideal training set. Kirkpatrick et al. [8] propose Elastic Weight Consolidation (*EWC*) to remember old tasks by constraining the important weight parameters to stay close to their old values. EWC builds a Fisher matrix using training data to determine which weights it should change. One limitation of this approach is that there will be conflicting constraints for the weight parameters which are shared between old tasks and new task. GANs have recently become popular for storing distribution of the data [13, 14]. Shin et al. [6] propose a novel framework - Deep Generative Replay with a cooperative dual model architecture consisting of a deep generative model and a task solving model to mitigate catastrophic forgetting. Kemker et al. [15] propose FearNet, a brain-inspired model for incremental learning which does not require storing old data. It consists of a dual memory system, where one is for short-term memories and other for long-term memory. Instead of learning the representation, they use pretrained ResNet [16]

embeddings to obtain features to be fed into their model. Our method also belongs to this category in that we do not require the old data. The key difference is that we encode the data firstly and store the encoded, low-dimensional data and decoder network for the consecutive tasks. Although we need to store the encoded exemplars, our experiments show that the encoded exemplars significantly improve the incremental learning performances compared to these prior techniques without using any form of old data at all.

3 IL4IoT Framework

In this section, we firstly define our problem and give its mathematical description. Then we describe the IL4IoT framework and its main components. Lastly, we explain the three key steps thoroughly and provide the algorithm details respectively.

3.1 Problem Statement

We focus on incremental learning for IoT devices. Just as Liu et al. [2] stated, we are given a sequence of supervised learning tasks $T = (T_1, T_2, \dots)$. Let us denote n_i be the total number of samples in each task T_i . And T_i corresponds to a labeled dataset $D_i = \{(x_{i,j}, y_{i,j}), j = 1, 2, \dots, n_i\}$, where x_{ij} is the j^{th} sample of task T_i and $y_{i,j}$ is its label. Our problem is class-incremental learning, in which each task has totally different classes. Assuming that Y_i and Y_k are the label sets for T_i and T_k respectively, for any y' belonging to Y_i and any y'' belonging to Y_k , there exists $y' \neq y''$. Then with the sequentially arriving tasks $T = (T_1, T_2, \dots)$, we need a mechanism to continually update the model to keep competitive in all of the arrived tasks. Considering that the limited computing power, storage, and memory of IoT devices, we want the framework to be as lightweight as possible.

3.2 Main Components

In order to retain the previously learned knowledge and assist the knowledge accumulation, in each stage, there are two cooperative parts: a knowledge base and a task-solving model (called as solver). In our IL4IoT framework, knowledge base consists of the encoded exemplars and their decoder for each task. So, with the task sequence, the knowledge base will be refined continually. Then based on the continually refined knowledge base in each stage, we can obtain the solver for each stage. As the framework shown in Fig. 2, it contains three key steps:

Step 1: Exemplars Selection. In order to reduce the computational complexity, we leverage autoencoder to learn hidden representation of the data, based on which we conduct exemplars selection with herding algorithm. See Sect. 3.3.

Step 2: Knowledge Refinement. In order to further reduce the computational complexity and protect the data privacy, we further apply autoencoder to learn representation of all the exemplars, and obtain the extracted knowledge including the encoded exemplars and their decoder, which will be retained and transferred to the following task. Based on the extracted knowledge of the former tasks, we reconstruct the encoded exemplars into exemplars with original dimensions. In order to further keep the storage requirements stable, we conduct exemplars reduction for the former tasks. And by combining the exemplars for former tasks with the exemplars for the newly arrived task, we obtain the exemplars for all the classes, based on which we achieve the knowledge refinement and obtain the newly refined knowledge. See Sect. 3.4.

Step 3: Solver Construction. Based on the refined knowledge, we accomplish the solver construction. See Sect. 3.5.

3.3 Step 1: Exemplars Selection

Assuming that the original training data is $D_i = \{(X_i, Y_i), i = 1, 2, \dots\}$, $X_i = \{x_{i,j}\}_{j=1}^{j=n_i}$, $Y_i = \{y_{i,j}\}_{j=1}^{j=n_i}$. In order to reduce the computational complexity, we apply the traditional autoencoder model [17] to map the original data $X_i \in [0, 1]^d$ to a hidden representation $Z_i \in [0, 1]^{d'}$ through the function $Z_i = \mathcal{F}_i(W_i X_i + b_i)$, parameterized by $\theta_i = \{W_i, b_i\}$. Then the hidden representation Z_i can be mapped back to a reconstructed vector $\widehat{X}_i \in [0, 1]^d$, where $\widehat{X}_i = \mathcal{G}_i(W'_i Z_i + b'_i)$ with parameter $\beta_i = \{W'_i, b'_i\}$. The parameters of \mathcal{F}_i and \mathcal{G}_i can be optimized to minimize the average reconstruction error:

$$\theta_i^*, \beta_i^* = \arg \min_{\theta_i, \beta_i} \frac{1}{n} \sum_{j=1}^{n_i} L(X_i, \widehat{X}_i) \quad (1)$$

where L is a loss function such as the mean squared error $L(X_i, \widehat{X}_i) = \|X_i - \widehat{X}_i\|_2$.

In order to alleviate catastrophic forgetting, we should retain the knowledge of the previous tasks. Since exemplars can approximately represent the whole training samples, we can regard the exemplars of each class as the knowledge of tasks and transfer them to the subsequent tasks. Furthermore, by considering the constrained resources of IoT devices, we assume Q is the total number of exemplars that can be stored. Then for each class, there are $q = Q/c$ exemplars where c is the number of classes already observed. Similar with the herding algorithm [18], we create a representative set of samples on behalf of the whole training samples. For any class k , samples can be selected to add to the exemplar set, which cause the average feature vector over all exemplars to best approximate the average feature vector over all training samples. Assuming that $\mu_i^{(k)}$ is the average feature vector over all the training samples of class k , which can be computed as follows:

$$\mu_i^{(k)} = \frac{1}{q} \sum_{j=1}^{n_i} z_{i,j} \mathbb{1}_{y_{i,j}=k} \quad (2)$$

where $\mathbb{1}_{y_{i,j}=k}$ is an indicator function which is used to choose the encoded samples of class k . The exemplars of class k in task T_i can be sequentially choose as follows:

$$r_{i,h}^{(k)} = \arg \min_{t \in [1, 2, \dots, n_i]} \|\mu_i^{(k)} - \frac{1}{h} [z_{i,t} \mathbb{1}_{y_{i,t}=k} + \sum_{j=1}^{h-1} z_{i,j} \mathbb{1}_{y_{i,j}=k}]\| \quad (3)$$

Combining the exemplars of all the classes in the task, we can obtain the exemplars set R_i for each task i . The detailed algorithm is listed in Algorithm 1.

Algorithm 1. IL4IoT SelectExemplars

Require: Training dataset of task i $D_i = \{(X_i, Y_i), i = 1, 2, \dots\}$, $X_i = \{x_{ij}\}_{j=1}^{j=n_i}$, $Y_i = \{y_{ij}\}_{j=1}^{j=n_i}$, the number of exemplars for each class already arrived: q .

Ensure: Exemplars R_i of task i .

- 1: Obtain \mathcal{F}_i and \mathcal{G}_i for X_i by equation (1);
 - 2: $Z_i \leftarrow \mathcal{F}_i(X_i)$;
 - 3: $L \leftarrow \text{distinct}(Y_i)$;
 - 4: **for** $k = 1, 2, \dots, \text{len}(L)$ **do**
 - 5: **for** $j = 1, 2, \dots, n_i$ **do**
 - 6: **if** $y_{ij} == L[k]$ **then**
 - 7: $A \leftarrow A \cup z_{ij}$;
 - 8: $m \leftarrow m + 1$;
 - 9: **end if**
 - 10: **end for**
 - 11: $\mu \leftarrow \frac{1}{m} \sum_{z \in A} z$;
 - 12: **for** $h = 1, 2, \dots, q$ **do**
 - 13: $r_{i,h}^{(k)} \leftarrow \arg \min_{z \in A} \|\mu - \frac{1}{h} [z + \sum_{t=1}^{h-1} z_{i,t}]\|$;
 - 14: $r_i^{(k)} \leftarrow \bigcup_{h=1}^q \mathcal{G}_i(r_{i,h}^{(k)})$;
 - 15: **end for**
 - 16: $R_i \leftarrow \bigcup_{k=1} r_i^{(k)}$.
 - 17: **end for**
-

3.4 Step 2: Knowledge Refinement

Since exemplars can approximately represent the whole training samples, if we want to retain the knowledge of former tasks, we can transfer the exemplars to the subsequent tasks. However, direct transfer of exemplars without any further operations will pose data privacy problem. Considering this problem, before the transfer, we extract knowledge from these exemplars not only to retain the knowledge, but also to protect the data privacy.

We use a shallow autoencoder \mathcal{F}_i to map the exemplars R_i to hidden representation, and further map back to reconstructed exemplars by another function \mathcal{G}_i . And the parameters of \mathcal{F}_i and \mathcal{G}_i can be optimized by using the Eq.(1).

Then the encoded exemplars $\mathcal{F}_i(R_i)$ and the function \mathcal{G}_i can form the extracted knowledge K_i for the already arrived i tasks, where $K_i = \{\mathcal{F}_i(R_i), \mathcal{G}_i\}$.

In order to alleviate catastrophic forgetting, we should retain the learned knowledge in all the former i tasks when we learn newly arrived task T_{i+1} . Then based on the transferred knowledge K_i , we can obtain the reconstructed exemplars \widehat{R}_i in original dimension by using the following equation:

$$\widehat{R}_i = \mathcal{G}_i(\mathcal{F}_i(R_i)) \quad (4)$$

As we know, the number of exemplars can be stored is fixed to Q , then before the exemplars combination with the exemplars of the newly arrived task T_{i+1} , which can be obtained by the exemplars selection step, we need to reduce the existing exemplars of all the former tasks. Since the exemplars set R_i is a prioritized list, where q is the target number of exemplars per class in the current training stage, we should only choose the first q exemplars and obtain the reduced exemplars set $\widetilde{R}_i = \widehat{R}_i[1, 2, \dots, q]$. Then by mixing the reduced exemplars \widetilde{R}_i for all the former tasks with the exemplars \overline{R}_{i+1} of current task, we can further obtain the exemplars for all the $i + 1$ tasks as follows:

$$R_{i+1} = \left(\bigcup_{j=1}^i \widetilde{R}_j \right) \bigcup R_{i+1} \quad (5)$$

Based on the refined exemplars R_{i+1} , we can refine the knowledge by using Eq. (1) and obtain the refined knowledge $K_{i+1} = \{\mathcal{F}_{i+1}(R_{i+1}), \mathcal{G}_{i+1}\}$. The detailed algorithm is listed in Algorithm 2.

3.5 Step 3: Solver Construction

Considering the constrained resources of IoT devices, we use a nearest-mean-of-exemplars classification strategy to accomplish the classification, which is also used in iCaRL [5]. Then in order to predict the label for a new sample x , we can compute the mean value $\alpha_i^{(k)}$ of the refined exemplars for each class k by using function \mathcal{G}_i .

$$\alpha_i^{(k)} = \frac{1}{q} \sum_{j=1}^{n_i} \mathcal{G}_i(z_{i,j}) \mathbb{1}_{y_{i,j}=k} \quad (6)$$

By comparing the similarity between the new input x with the mean value of each class, we can assign the class label with highest similarity:

$$y^* = \arg \min_{k \in [1, 2, \dots, c]} \|x - \alpha_i^{(k)}\|. \quad (7)$$

The solving model can be achieved by the algorithm listed in Algorithm 3.

Algorithm 2. IL4IoT RefineKnowledge

Require: Exemplars of task i : R_i , total number of exemplars: Q , number of classes before task $i + 1$: c , target number of exemplars per class in stage $i + 1$: q .

Ensure: Knowledge base in session $i + 1$: K_{i+1} .

- 1: Obtain \mathcal{F}_i and \mathcal{G}_i for R_i by equation (1);
 - 2: $K_i \leftarrow \{\mathcal{F}_i(R_i), \mathcal{G}_i\}$;
 - 3: $\widetilde{R}_i \leftarrow \mathcal{G}_i(\mathcal{F}_i(R_i))$;
 - 4: $\widetilde{R}_i \leftarrow \bigcup_{j=1}^c \widetilde{R}_i[\frac{Q}{c}(j-1)+1 : \frac{Q}{c}(j-1)+q]$;
 - 5: $\overline{R}_{i+1} \leftarrow \text{SelectExemplars}(D_{i+1})$;
 - 6: $R_{i+1} \leftarrow \overline{R}_{i+1} \cup \widetilde{R}_i$;
 - 7: Obtain \mathcal{F}_{i+1} and \mathcal{G}_{i+1} for R_{i+1} by equation (1);
 - 8: $K_{i+1} \leftarrow \{\mathcal{F}_{i+1}(R_{i+1}), \mathcal{G}_{i+1}\}$.
-

Algorithm 3. IL4IoT ConstructSolver

Require: New input data: x , Knowledge base in session i : $K_i = \{\mathcal{F}_i(R_j), \mathcal{G}_i\}$, the number of exemplars for each class already arrived: q , the number of classes: c

Ensure: Class label k^* .

- 1: **for** $k = 1, 2, \dots, c$ **do**
 - 2: $\alpha_i^{(k)} = \frac{1}{q} \sum_{j=1}^q \mathcal{G}_i(\mathcal{F}_i(R_j))$
 - 3: **end for**
 - 4: $k^* \leftarrow \arg \min_{k \in [1, 2, \dots, c]} \|x - \alpha_i^{(k)}\|$
-

4 Experiments

In this section, we conduct experiments on two benchmark datasets to compare the performance of our IL4IoT method with three other methods under the incremental learning settings. Firstly, we introduce the MNIST and SDA datasets and their incremental learning settings. Then we illustrate the comparative analysis of our IL4IoT method and the three other methods.

4.1 Datasets

MNIST dataset [19]: this dataset consists of 60,000 images of handwritten digits from 0 to 9. To simulate incremental learning, we divide the MNIST dataset into several subsets of classes. Each subset is a task. In our experiments, we divide the MNIST dataset into five tasks (2 classes per task). Each task consists of two digits (classes) respectively, such as $\{0, 1\}$, $\{2, 3\}$, $\{4, 5\}$, $\{6, 7\}$ and $\{8, 9\}$. Then the aim of our system is to learn these five subsets as five tasks in a incremental fashion.

SDA dataset [20]: this dataset consists of 285,000 samples, which is obtained by the mean values and variations of 3-axis accelerometer, a 3-axis gyroscope, and a 3-axis magnetometer. Each sample corresponds to a daily and sport activity. There are nineteen activities including sitting, standing, lying on back, lying on right side and so on. In our experiments, we divide the SDA dataset into ten tasks (2 classes for the first 9 tasks and 1 class for the last one task), five tasks (4 classes for the first 4 tasks and 3 classes for the last one task).

4.2 Training Details

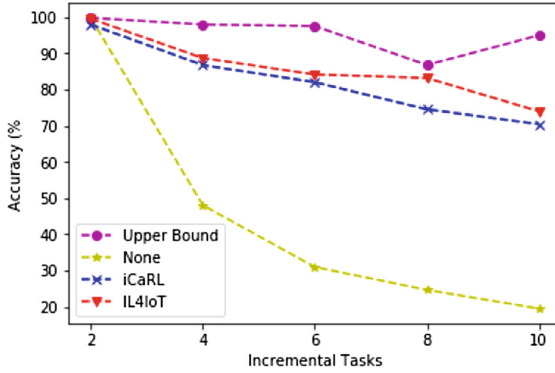
We compare the performance of our IL4IoT method with three other methods. Our method is denoted in the figure as **IL4IoT**. The iCaRL method is denoted in the figure as **iCaRL**. We specify the upper bound (denoted as **UB**) by assuming a situation when the model can be trained on all of the training data. A baseline of naively trained model, i.e., where catastrophic forgetting exists yet without any effort to alleviate it, is denoted as **None**. All the experiments are conducted on a Mac Pro with 3.1 GHz Dual-Core Intel I5 processor and 8 GB memory.

On MNIST dataset, we adopt a 2-layer convolutional network (CNN) with $5 * 5$ convolutions and 32, 64 filters for each layer. And we set the basic unit size of fully connected layer to 1024 in UB experiments, None experiments, and iCaRL experiments. The exemplars size is set to 120 in iCaRL and IL4IoT experiments. The basic unit size of hidden layer in our IL4IoT’s autoencoder is 64. On SDA dataset, we adopt a 3-layer network (basic unit size is 30) in UB experiments, None experiments, and iCaRL experiments. The exemplars size is set to 240 in iCaRL and IL4IoT experiments. The basic unit size of hidden layer in our IL4IoT’s autoencoder is 10.

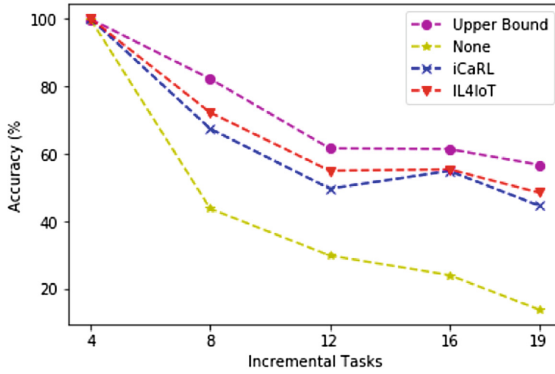
4.3 Comparative Analysis

Figure 3 shows the classification accuracies of each method for each task as the tasks are incrementally learned on MNIST dataset and SDA dataset. As shown in Fig. 3(a), on MNIST dataset, if we do not alleviate catastrophic forgetting, the accuracy for the incremental tasks will decline sharply with the incremental tasks. When it comes to the last training session, the accuracy will decline to 19.37%. In contrast, if we can train on all the training data, in every training session, we can obtain the upper bound accuracies, which are all higher than 95%. When we use iCaRL to alleviate catastrophic forgetting, the accuracies for the incremental tasks have a small amount of reduction from the upper bound accuracies, which are 97.07%, 87.80%, 83.00%, 74.27%, and 69.04% respectively. Compared with iCaRL, our IL4IoT can obtain the accuracies of 99.72%, 91.77%, 84.84%, 82.94%, and 76.58%, which show about 4.93% improvements on average. And the results demonstrate that although our method does not leverage the knowledge distillation technique to refine the representation of the samples and only uses the encoded exemplars, it can achieve higher accuracy than the iCaRL method.

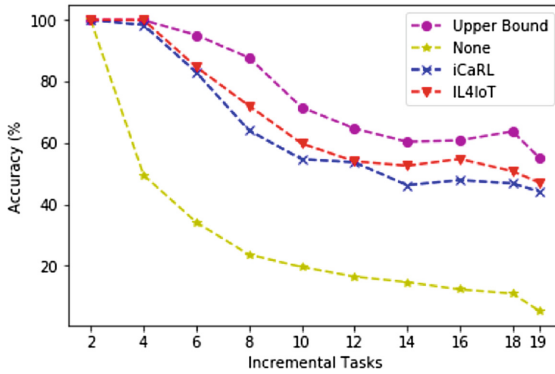
As to the classification accuracies on SDA dataset, similar results can be observed when the number of incremental tasks are five (4 classes for the first 4 tasks and 3 class for the last one task). As shown in Fig. 3(b), if we do not alleviate catastrophic forgetting, the accuracies will decline sharply from 99.87% to 14.01%, which demonstrates that when catastrophic forgetting is not considered, the model will be ineffective and non-applicable for the incremental tasks. In contrast, in upper bound experiments, the accuracies of the incremental tasks are four times higher than that of None experiments at most. In iCaRL, the accuracies increases about 21% than None experiments and the average accuracy is



(a) MNIST Dataset (5 incremental tasks)



(b) SDA Dataset (5 incremental tasks)



(c) SDA Dataset (10 incremental tasks)

Fig. 3. Comparisons of accuracies between None, iCaRL, IL4IoT, and Upper Bound on MNIST dataset with 5 incremental tasks, SDA dataset with 5 and 10 incremental tasks

about 63.39%. In our IL4IoT, the accuracies can achieve about 3% improvements over the accuracy of iCaRL, which demonstrates our IL4IoT is more effective than the iCaRL method.

As shown in Fig. 3(c), in upper bound experiments, the accuracies of the incremental tasks are from 99.99% to 57.54% and if the number of incremental tasks is four, the accuracies can be higher than 88.2%. However, if we do not alleviate catastrophic forgetting, the accuracies will decline sharply when the number of tasks comes to four and the accuracy will decline to 23.3%, which all demonstrate the huge effect of catastrophic forgetting. In iCaRL experiments, the accuracies have about 35.5% improvements from the None experiment on average. When the number of tasks is four, the accuracy is 67.4%. In our IL4IoT experiments, the accuracies have about 39.1% and 3.6% improvement from the None method and the iCaRL method on average. When the number of tasks is four, the accuracy is 74.3%, which is 8.4% improvement over the iCaRL method.

We also conduct experiments on running time of each method on the MNIST dataset and the SDA dataset. As shown in Table 1, in None experiment, the running times stay stable with the incremental tasks, which will cost 46s on average. In contrast, in upper bound experiment, since we need to train on all the training data, the running times will increase greatly with the incremental tasks, which is 153s on average. In iCaRL, since we only leverage the exemplars of former tasks, the running times keep stable and the average value drops to 58s. In our IL4IoT, since we leverage the encoded exemplars and do not leverage neural network classification method, the running times of our method are lower than iCaRL and UB, which is 41 s.

As to the running times on the SDA dataset, similar results can be observed when the total number of incremental tasks are five (4 classes for the first 4 tasks and 3 class for the last one task). As shown in Table 2, in upper bound experiments, the running times increase sharply when it comes to the fifth task. The running times will increase to 803.5s, which is not acceptable in highly dynamic IoT environments. The running times of our method are 2.2 times and 2.4 times less than that of None method and iCaRL method, which demonstrate the efficiency of our method. Similar results can be observed when the number of incremental tasks are ten (2 classes for the first 9 tasks and 1 class for the last one task) shown in Table 3. The running times of our method are 1.6 times, 2.2 times, and 10.1 times shorter than that of None method, iCaRL method, and upper bound method.

The experimental results demonstrate that our method can achieve with higher accuracy and higher efficiency. Since autoencoder can learn effective latent representation of the original data, even with the encoded exemplars, the accuracy of our IL4IoT method is higher than the iCaRL method's. Besides, in our IL4IoT method, we also conduct representation learning to increase accuracy. However, the difference is that we leverage autoencoder to refine the representation continually, which is a more efficient strategy than the knowledge distillation used in iCaRL method. The reasons for the higher efficiency are two aspects. One is that the exemplars selection and the solver construction in encoded form

Table 1. Comparisons of running time between None, iCaRL, IL4IoT, and Upper Bound on MNIST dataset with 5 incremental tasks

Task i	None	iCaRL	IL4IoT	Upper Bound
i = 1	44.14	49.29	45.71	50.20
i = 2	43.66	60.87	42.67	91.13
i = 3	43.84	55.53	38.40	183.93
i = 4	49.83	60.97	38.83	193.18
i = 5	50.99	63.13	40.61	248.87

Table 2. Comparisons of running time between None, iCaRL, IL4IoT, and Upper Bound on SDA dataset with 5 incremental tasks

Task i	None	iCaRL	IL4IoT	Upper Bound
i = 1	78.80	45.66	28.40	79.10
i = 2	79.24	95.24	34.16	157.76
i = 3	77.71	99.56	34.64	235.26
i = 4	78.85	96.72	38.84	585.81
i = 5	61.33	77.64	34.18	803.51

Table 3. Comparisons of running time between None, iCaRL, IL4IoT, and Upper Bound on SDA dataset with 10 incremental tasks

Task i	None	iCaRL	IL4IoT	Upper Bound
i = 1	41.75	12.72	26.32	40.61
i = 2	38.98	46.23	20.99	78.45
i = 3	39.64	48.57	18.02	118.87
i = 4	40.13	46.92	16.96	162.10
i = 5	39.22	46.35	17.57	194.35
i = 6	40.61	47.29	19.17	233.49
i = 7	41.26	48.81	20.99	272.50
i = 8	41.51	50.72	23.62	310.68
i = 9	42.62	52.30	26.12	348.54
i = 10	20.06	32.79	21.43	366.67

will obviously reduce the computational time. The other is that the solver construction strategy is more efficient than the classic neural network classification.

As to the data privacy issue, we analyze the main steps of all of the above methods - None, UB, iCaRL and IL4IoT. In None, we don't consider the catastrophic forgetting problem and train model only on the training data of the newly arrived classes. This method won't pose any data privacy problem. In UB, due to that all of the training data of already arrived classes can be accessed, huge

data privacy problem will be caused. In iCaRL, although only exemplars will be stored, the storage of the original form of exemplars will pose data privacy problem. In our IL4IoT method, we conduct knowledge extraction based on exemplars and only retain the encoded exemplars and decoder network, which are not related to the data privacy. In conclusion, IL4IoT compared to iCaRL and UB, can improve the accuracy and computing-resource efficiency, without sacrificing privacy-protection. All these characteristics make our IL4IoT more suitable for resource-constrained IoT devices for incremental learning.

5 Discussion and Future Work

The challenge addressed in this paper is how to continually learn new concepts from data stream without catastrophic forgetting, especially for resource-constrained IoT devices. The methods aiming to alleviating with catastrophic forgetting can be divided into two categories: requiring old data or not requiring old data. As we know, the methods require part of all of the old data significantly improve the incremental learning performances compared to the techniques without using any form of old data at all. However, directly storing part or all of the old data not only costs huge computational resources and memory storage, but also it will pose data privacy problem. Therefore, deeply extracting knowledge from the old data to support the incremental learning is considered to be a reasonable and promising method.

Although the experiments on MNIST dataset and SDA dataset demonstrate the effectiveness and efficiency of our IL4IoT approach, we need to further test our method on other large datasets such as CIFAR [21], ImageNet ILSVRC 2012 dataset [22]. There are other open issues to discuss. In our experiments, we only use the basic autoencoder, not considering other sophisticated autoencoders, such as variational autoencoder [23], sparse autoencoder [24], and denoising autoencoder [25]. Will these more sophisticated autoencoders make the accuracy increase more effectively? Also, IL4IoT is still lower than the upper bound on accuracy, which is obtained by training on all training samples of all classes. We plan to analyze the reasons for this in more detail. The problems discussed above are the future directions we will focus on.

6 Conclusion

In this paper, we propose the IL4IoT, a lightweight incremental-learning framework for IoT devices, which consists of two cooperative parts: a continually updated knowledge base and a task-solving model. Through this framework, we can achieve incremental learning while alleviating the catastrophic forgetting issue, protecting privacy, and reducing computational requirements. Our experimental analysis demonstrates promising results compared to the best available solutions with an overall accuracy improvement of 3.83% and a computational cost reduction of 47% using two benchmark datasets. In conclusion, our IL4IoT method not only can achieve higher accuracy than iCaRL with shorter cost

times, but also can effectively protect the data privacy, which all demonstrate that our IL4IoT is more suitable for resource-constrained IoT devices for incremental learning.

References

1. Wu, Q., et al.: Cognitive internet of things: a new paradigm beyond connection. *IEEE Internet Things J.* **1**(2), 129–143 (2014). 2014
2. Liu, B.: Lifelong machine learning: a paradigm for continuous learning. *Front. Comput. Sci.* **11**(3), 359–361 (2017)
3. Gupta, C., et al.: ProtoNN: compressed and accurate kNN for resource-scarce devices. In: *ICML* (2017)
4. Kumar, A., Goyal, S., Varma, M.: Resource-efficient machine learning in 2 KB RAM for the internet of things. In: *ICML* (2017)
5. Rebuffi, S.A., Kolesnikov, A., Lampert, C.H.: iCaRL: incremental classifier and representation learning. In: *CVPR* (2017)
6. Shin, H., Lee, J.K., Kim, J., Kim, J.: Continual learning with deep generative replay. In: *NIPS*, pp. 2990–2999 (2017)
7. Li, Z., Hoiem, D.: Learning without forgetting. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016*. LNCS, vol. 9908, pp. 614–629. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_37
8. Kirkpatrick, J., et al.: Overcoming catastrophic forgetting in neural networks. In: *Proceedings of the National Academy of Sciences*, p. 201611835 (2017)
9. Lopez-Paz, D., et al.: Gradient episodic memory for continual learning. In: *NIPS*, pp. 6470–6479 (2017)
10. Lane, N.D., Bhattacharya, S., Georgiev, P., Forlivesi, C., Kawsar, F.: An early resource characterization of deep learning on wearables, smartphones and internet-of-things devices. In: *IoT-App* (2015)
11. McCloskey, M., Cohen, N.J.: Catastrophic interference in connectionist networks: the sequential learning problem. *Psychol. Learn. Motiv.* **24**, 109–165 (1989)
12. Robins, A.: Catastrophic forgetting, rehearsal and pseudo rehearsal. *Connect. Sci.* **7**(2), 123–146 (1995)
13. Goodfellow, I., et al.: Generative adversarial nets. In: *NIPS*, pp. 2672–2680 (2014)
14. Mirza, M., Osindero, S.: Conditional generative adversarial nets. In: *Deep Learning Workshop NIPS* (2014)
15. Kemker, R., Kanan, C.: FearNet: brain-inspired model for incremental learning. In: *ICLR* (2018)
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *CVPR*, pp. 770–778 (2016)
17. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. *Science* **313**(5786), 504 (2006)
18. Welling, M.: Herding dynamical weights to learn. In: *ICML* (2009)
19. LeCun, Y., Cortes, C., Burges, C.: MNIST handwritten digit database
20. Barshan, B., Yükek, M.C.: Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. *Comput. J.* **57**(11), 1649–1667 (2014)
21. Krizhevsky, A., Hinton, G.: Learning multiple layers of features from tiny images (2009)

22. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis. (IJCV)* **115**(3), 211–252 (2015)
23. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: *ICLR* (2014)
24. Makhzani, A., Frey, B.: K-sparse autoencoder. In: *ICLR* (2014)
25. Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.: Extracting and composing robust features with denoising autoencoders. In: *ICML* (2008)



Enhanced Buying Experiences in Smart Cities: The SMARTBUY Approach

Lorena Bourg¹, Thomas Chatzidimitris^{2,8}, Ioannis Chatzigiannakis^{3,8},
Damianos Gavalas^{4,8(✉)}, Kalliopi Giannakopoulou^{5,8},
Vlasios Kasapakis^{2,8}, Charalampos Konstantopoulos^{6,8},
Damianos Kyriadis^{6,8}, Grammati Pantziou^{7,8},
and Christos Zaroliagis^{5,8}

¹ Planet Media Studios, Madrid, Spain
lbourg@gmail.com

² Department of Cultural Technology and Communication,
University of the Aegean, Mytilene, Greece
{tchatz, v.kasapakis}@aegean.gr

³ Department of Computer, Control and Informatics Engineering,
Sapienza University of Rome, Rome, Italy
ichatz@diag.uniroma1.it

⁴ Department of Product and Systems Design Engineering,
University of the Aegean, Syros, Greece
dgavalas@aegean.gr

⁵ Department of Computer Engineering and Informatics, University of Patras,
Patras, Greece
{gianakok, zaro}@ceid.upatras.gr

⁶ Department of Informatics, University of Piraeus, Piraeus, Greece
{konstant, dkypriad}@unipi.gr

⁷ Department of Informatics and Computer Engineering,
University of West Attica, Athens, Greece
pantziou@uniwa.gr

⁸ Computer Technology Institute and Press (CTI), Patras, Greece

Abstract. The establishment of shopping malls and the growth of online shopping increasingly diminishes the turnover of “small”, independent retailers in urban environments. However, retailers could reverse this trend through complementing the offline experiences they already offer with online offerings and establishing business “alliances” to achieve economies of scale and enable the provision of innovative digital services. The EU-funded project SMARTBUY aims at realizing the concept of a “distributed shopping mall” ecosystem which allows retailers to band together in a large commercial coalition which generates added-value for its retailers-members and customers: centralized products and services inventory management; geo-located marketing of products/services; location-based search for products offered by nearby retailers; personalized recommendations for purchasing products based on innovative recommendation systems. In effect, SMARTBUY proposes a blended shopping paradigm, wherein the benefits of online shopping are combined with the appeal

of traditional store shopping. The article provides an overview of the main outcomes and achievements of SMARTBUY. It also reports on conclusions drawn in the context of the project's official pilot execution in four European cities.

Keywords: E-commerce · Retailer · Shopping · Inventory management · Product · Service · Smart cities · Smart retailing · Geo-located marketing · Location-based search · Recommendation

1 Introduction

The decline of isolated retailers and the rise of shopping malls have been major trends in retailing for decades, worldwide. Retail agglomeration in shopping malls has been found useful to both consumers as well as firms. Shopping mall agglomeration offers the benefits of higher footfalls, shared -physical and digital- infrastructure, cooperative marketing plans and high impulse purchases [17].

The fast-paced development of ICTs has also dramatically changed the way people shop in the past 15 years. Besides shopping at physical stores, with the aid of ICT, consumers are able to shop via the Internet. This new type of shopping mode (e-shopping or online shopping) frees consumers from having to physically visit stores and puts extra competitive pressure to local small retailers since customers often opt to shop online due to the attained cost and -travel- time savings [8, 16].

Nevertheless, the above discussed trends could presumably be reversed if retailers were to integrate ICT services into the offline experiences they already offer and establish business “alliances” to achieve economies of scale and enable the provision of innovative services.

Along this line, the H2020 SMARTBUY project¹ (Jan 2016–Jan 2019) aims at realizing the concept of a “distributed shopping mall” ecosystem which allows retailers to band together in a large commercial coalition which generates added-value for its retailers-members and customers: centralized products and services inventory management; geo-located marketing of products/services; location-based search for products offered by nearby retailers; personalized recommendations for purchasing products based on innovative recommendation systems. From the customer perspective, such an approach would combine the best of the two worlds: one the one hand, the benefits of online shopping (comparison of product prices across several stores, have guarantee on the availability of searched products, etc.); on the other hand, the multifarious appeal of traditional store shopping (social interaction, entertainment, movement, trip chaining), live in-store information, offline retailer loyalty, low product performance risk, zero delay for product delivery, etc. [6].

This article provides an overview of the main outcomes, contributions and findings of SMARTBUY. It also reports on experiences gained in the context of the project's official pilot execution in four European cities with diverse scale and retail market characteristics.

¹ <http://smartbuy.tech/>.

The remainder of the article is structured as follows: Sect. 2 reviews research related with the work presented herein. Section 3 discusses the scope of the project as well as its main scientific, technical and business objectives. Section 4 provides an overview of the system architecture of SMARTBUY and describes the functionality and interoperation among its key modules. Section 5 analyses results yield during the official pilot tests of the project. Finally, Sect. 6 concludes the paper.

2 Related Work

The six most-common indicators of smart cities are smart economy, smart people, smart governance, smart mobility, smart environment, and smart living [7]. Smart economy refers to technologies supporting new forms of collaboration and value creation that lead to innovation, entrepreneurship and competitiveness [10]. In this context, Pantano and Timmermans in their seminal work on smart retailing, emphasize that “the emerging idea of smart retailing would reflect a particular idea of retailing, where firms and consumers use technology to reinvent and reinforce their role in the new service economy, by improving the quality of their shopping experiences” [13].

The smart retailing culture is characterized by the extensive use of mobile technologies, high connectivity, ubiquitous computing and contactless technologies, which enable consumers to experience shopping differently [14]. Thus, it is not surprising that there is an increasing awareness in marketing about the need to develop new mobile retailing strategies. Initial efforts involved the usage of SMS for advertising purposes. Later, it expanded to mobile apps, which allow consumers to easily find, compare and order products, access news on products and services, create shopping lists, locate products and stores, etc. [14]. More recently, it encompassed smart stores with iBeacon infrastructures used to track the exact customer location and push product-related notifications, as well as NFC-equipped smartphones executing augmented reality apps to add virtual elements (textual information, images, videos, 3D animation clips, etc.) to products of interest when visiting the actual store [15].

Among other aspects, knowledge management has been recognized as a critical success factor for (mobile) smart retailing, since data collection and consumer analysis can be facilitated by the adoption of new technologies [11]. Mobile/ubiquitous computing provide the required technological infrastructure to collect valuable knowledge from/for consumers (e.g., by codifying habits, processing clients’ transaction information and identifying changes in consumers’ behavior for supporting retailers, etc.), and feed knowledge into the service (i.e., by providing enriched information about certain products, or drawing the attention of consumers to -otherwise missed- products of interest) [1, 2]. Along this line, mobile recommender systems (RS) enable the adaptation of product information to make it more relevant to consumers’ information needs [4, 5, 19]. Mobile RSs may serve both in-store and online purchase situations as they “elicit the interests or preferences of individual users for products either explicitly or implicitly, and make recommendations accordingly” [9].

For instance, Yang et al. proposed a location-aware RS that correlates customers’ shopping needs with location-dependent vendor offers and promotions [25]. Yuan and Tsao introduced a framework which enables the creation of tailor-made campaigns

targeting users according to their location, needs and devices' profile (i.e. contextualized mobile advertising) [26]. More recently, Chen and Ji proposed an approach for smart advertising in smart cities, with the use of crowdsourced trails provided by mobile and wearable IoT devices [3]. The authors argue that, as user behavior is tracked via a device, analysis of these trails generates important clues for market planning.

SMARTBUY builds upon these developments in mobile retailing proposing the involvement of retailers and clients in a sort of “smart-partnership”. This approach provides the basis for value creation via the implementation of smart marketing and commerce services which will be analyzed in the sequel.

3 Project Scope and Objectives

SMEs lack visibility in the digital world. A large percentage of small businesses in Europe do not have a website; only a few have their product catalogues on-line, while most rely on social media marketing to disseminate promotional offers. They are rarely visible on large commercial platforms. In addition, the use of commercial platforms is very costly for SMEs. On the other hand, many customers would be interested in supporting their local retail SMEs and buying from them instead of from popular e-commerce platforms. However, this is not straightforward in the absence of on-line platforms specifically dedicated to SMEs.

For small retailers who wish exposure of their products online, it would make sense to cooperate together at local level and create a local online catalogue of products where they could advertise to local consumers. Individual SMEs may lack necessary IT skills and visibility online, but if such a platform was created it would benefit all the small retailers involved. Nowadays, retail SMEs should obtain help to create such platforms at the local level. In particular, help linked to IT skills and marketing would be necessary.

SMARTBUY contributes in closing this gap providing the technological infrastructure for small and medium sized retailers to become *the place* where they can easily make their products and promotions visible on-line for local audiences at the right time and at the right moment; just in their smartphone and providing them also with competitive advantages of e-commerce: comparison of prices, choice of providers, reviews and specification awareness, guarantee of availability, etc. SMARTBUY provides SMEs retailers with an integrated suite of services allowing them to apply the tactics typically practiced by big retailers, thus offering an integrated digital and physical buying experience to customers.

SMARTBUY is thought as a service that achieves its bigger potential when applied in a delimited geographical area with dense deployment of commercial stores such as a city center. The service provides a location-based e-commerce (*l-commerce*) integrated infrastructure for *all* the small retailers in a geographical area to use. The *l-commerce* infrastructure allows conducting centralized searches of products provided by the

stores. Customers can have access to real-time information of the products, prices, availability, etc., with the convenience of being able to physical purchase in a local, near-to-home store, thus saving costs and delivery delays due to shipment.

SMARTBUY turns the physical stores of smart cities to a smart geographically distributed mall by providing the logical consistency needed for conducting centralized searches in heterogeneous and geographically distributed physical stores.

Figure 1 shows how potential customers with their smartphones are able, using the SMARTBUY system, to receive real time information with respect to the availability of the products that they wish to purchase from nearby stores. This awareness can be achieved through an explicit search on the SMARTBUY’s retailer’s stores databases but also through available location-based services (see Sect. 4.1).

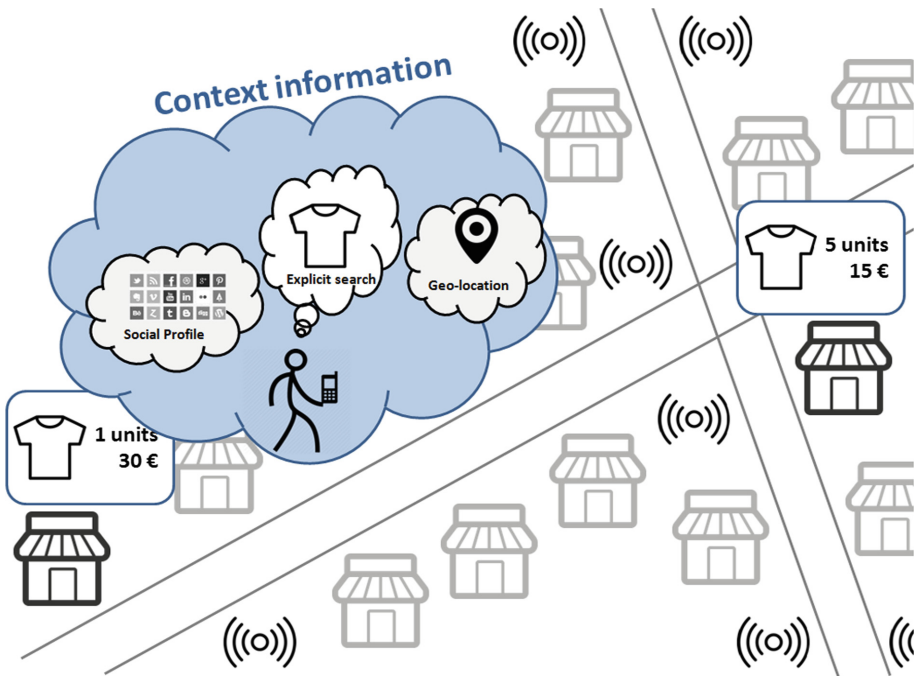


Fig. 1. Real-time centralized search for products offered by SME retailers’ stores.

SMARTBUY service is based on mature communication and information infrastructure components provided by the partners of the consortium. Nevertheless, a large-scale demonstration of the service is needed in order to validate the hypothesis on the impact that the service may have on the increase of sales and competitiveness of the small retailers.

SMARTBUY project pursues the following main scientific, technical and business objectives:

- Provision of services and functionalities targeted to two different user types, SME retailers and potential customers.

Services provision to potential customers allowing to:

- Conduct geo-located, centralized and ubiquitous searches on the inventory of services and products of member-stores of SMARTBUY.
- Check the availability, characteristics, price and reviews by third-parties of the products and services of their interest.
- Reserve the selected products and/or services online if the store had previously enabled this option.
- Review and share the acquired product and/or service online.
- Receive personalized recommendations and offers about products and services sold on nearby stores based on several contextual factors.

Services provision to retail SMEs stores allowing to:

- Integrate their services and products databases to an integrated geo-localized sales platform, attracting local audiences and increasing sales.
- Keep track of their inventory online.
- Enable online visibility for their products' catalogue and sales at low cost.
- Offer personalized and geo-located information and promotions to potential customers.
- Provide customized information about their products and services
- Provide real-time information about the availability of their products and services.
- Provision of a customized multichannel system for the interaction of potential customers with the platform (web portal and mobile app which takes advantage of geolocation capabilities).
- Large scale pilot experimentation and validation of the SMARTBUY platform (especially wrt sales and profit increase as well as the ability to reach new potential customers) in several urban environments with diverse geographical, cultural and market characteristics.

4 Overview of the SMARTBUY Architecture and Services

Figure 2 illustrates the high-level architecture of the SMARTBUY ecosystem. The system is built upon a cloud infrastructure (addressing security, privacy, scalability and interoperability requirements) on top of which several modules have been deployed. In the sequel, we describe the different modules integrated in the SMARTBUY system, emphasizing on their innovative aspects in relation to the state of the art.

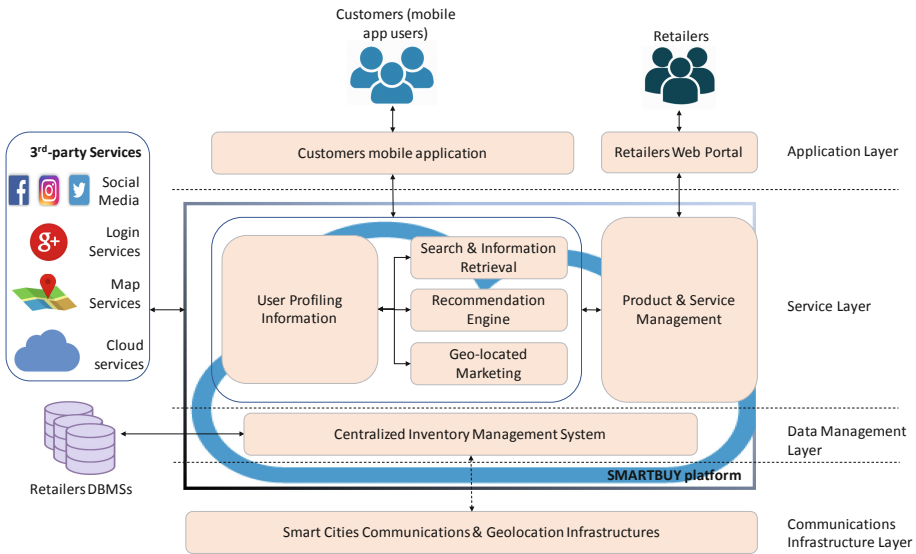


Fig. 2. System architecture of SMARTBUY.

4.1 Geo-located Marketing

Due to the ubiquity of mobile devices, the collection of large-scale, longitudinal data about human mobility is now commonplace. Medium-to-high resolution mobility data of individuals and signals of entire social systems can be captured from systems designed to enable communication and connectivity, such as cellular networks or WiFi APs (access points) [18]. Mobility traces are highly unique and identify individuals with substantial accuracy. Sensitive features can be extracted from mobility data, including home and work locations, visited places, or personality traits [18].

In the context of SMARTBUY, we hypothesize that human mobility data inferred from WiFi APs could boost the efficacy of marketing campaigns pursued by individual retailer-members or by the entire system as a whole [21]. The exploitation of the SMARTBUY geo-marketing tools requires that the APs of the retailer-members are configured so as to connect to the CityPassenger's CityScope gateway platform². The motivation behind using CityScope is that small retailers cannot handle the cost of hiring a full-time network administrator. With CityScope, management tasks are simplified and integrated into a unique tool that provides multiple functionalities: data analytics, management of connected devices, security, control, information exchange between remote APs, etc.

The key idea behind SMARTBUY's geo-marketing is to use the WiFi APs installed on retailers' premises as customer proximity detectors. The APs are able to capture the presence of users whose devices can be either registered or unregistered in the wireless network. The former connect to the AP (either automatically or manually through the

² <http://90.83.46.3/>.

SMARTBUY-branded WiFi captive portal (it is noted that the captive portal can be exploited to promote the SMARTBUY brand to non-registered consumers and encourage them to discover the services offered). The latter are detected through the probe request frames which are periodically transmitted from the WiFi radios of their smartphones (the unique MAC address of the radio transceiver is indicated on the header of those frames and may be used to identify users). The retailers APs are configured so as to report captured WiFi signals to CityScope. Among others, CityScope enables the information exchange among SMARTBUY-branded APs, even if they belong to different LANs and different retailers, effectively forming a wireless geo-marketing APs cluster at a city level. It is noted that the methods used for the collection, anonymization, and storage of data are GDPR-compliant.

This city-scale customer mobility inference infrastructure offers numerous advantages to the retailers. The first one is to be informed when a registered consumer is around, not only in the vicinity of their own store but also when connected to another SMARTBUY hotspot in the broader area. The user may then be sent notifications for personalized offers and discounts directly to her smartphone or be invited to install the mobile application and engage with the SMARTBUY community. The second one is to yield statistics about the connection of anonymous devices (i.e. non-SMARTBUY-registered) based on the logs systematically recorded by APs. These statistics are important to gain insights about the evolution of physical store traffic. At a higher level, these data can be used to track the behavior of anonymous consumers (trajectories, time spent at each location, etc.) in an area with several deployed WiFi APs. User mobility data captured by the geo-located marketing module are also exploited by the platform's recommendation engine in order to deliver relevant (i.e. user location history-informed) product recommendations to the customers.

4.2 Product and Services Data Management

The main interface among the retailers and the SMARTBUY system is the retailers' web portal. The latter is a web interface which offers the following key functionalities [22]:

- User management: registration and authentication of users (store managers)
- Creation/editing of store (retail) details, like title, location, description, contacts, opening hours, etc.
- Creation/editing of catalogues.
- Creation/editing of products: name, category/subcategory, attributes (e.g. color, weight, size), short/long description, price, stock, photographs.
- Product promotions management: product, description, discount, start/end activation date.
- Statistics, like most viewed products, origin (location) of visitors, peak hours in the area of the store.

The retailers portal adopts responsive design principles; hence, it may be accessed from PC, tablet or smartphone.

4.3 Search and Information Retrieval

This core module enables end-users (customers) to perform searches for products and services offered in nearby retailer stores (*l*-commerce) over a unified, centralized inventory of products administered by the geographically distributed retailer-members of the SMARTBUY community. The module combines the search results with personalized product promotions and recommendations.

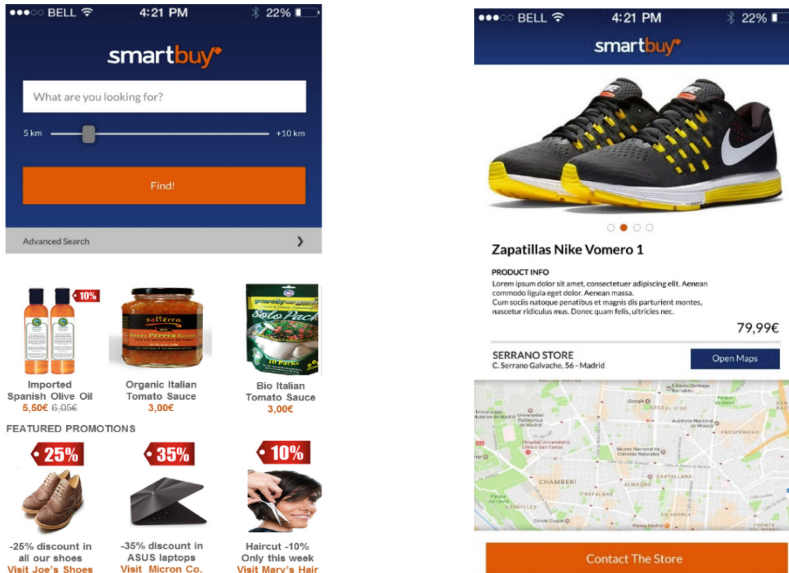


Fig. 3. Representative screens of the SMARTBUY app.

The module is accessed via the SMARTBUY native mobile application (see Fig. 3), which offers the following key functionalities [20]:

- Location-based search for products and services, returning the list of products and stores which fulfill the search criteria.
- Application of several filtering options (maximum distance of retailers' location from the user's current location, product categories, price range) on location-based searches.
- Browsing of product details (title and description, attributes like price, size and color, availability) and store details (location, contact information, etc.).
- Adding products to a 'wishlist' or 'sharing' them with others.
- Reserving a product, to ensure its availability, for a limited period of time.

- Viewing promotional offers managed by the retailers as well as personalized lists of recommended products which are relevant to the customers' consumer profile and mobility pattern (see Sect. 4.4).

4.4 Recommendation Engine

The purpose of the SMARTBUY Recommendation Engine (RE) is to derive meaningful product recommendations addressed to mobile SMARTBUY users, so as to draw the attention of users to products they would potentially be interested in, but they would not discover themselves alone. The core RE algorithm of SMARTBUY employs a collaborative filtering (CF)³ approach [5]. The main research contribution of our work lies in that, among other contextual factors (like current location, time, season, consumer behavior), the SMARTBUY's RE also takes into account the *location history* of the target user for delivering more meaningful recommendations. In particular, location history is exploited in two ways: (a) to attain good matchings among pairs of users (intuitively, users who move around the same neighborhoods or visit shops of the same category will have interests or restrictions in common); (b) to recommended products sold on retailer stores relatively far from the user's current location, provided that the stores are located in areas frequently visited by the user. Moreover, interfacing with social media (Facebook) has been implemented in order to access basic user demographic information and be able to derive relatively accurate recommendations even when user ratings data are relatively sparse [20].

The SMARTBUY mobile applications record in-app events (that is, events generated when users interact with the application, like keywords typed for product searches, product views, etc.). Effectively, the events are 'translated' to product ratings (e.g., when a user adds a product to her wishlist, this event is treated as a positive rating). The in-app events are timestamped (to capture the time context) and geo-stamped (to capture the mobility patterns of users, essentially, the GPS locations where they use the SMARTBUY app). Those data are uploaded, upon network availability, to the SMARTBUY's backend system (a web service endpoint has been implemented for that purpose) and stored in a database. SMARTBUY also makes use of WiFi APs deployed in the premises of retailers to accurately track user trajectories and record the number of users' 'physical' visits to stores (as explained in Sect. 4.1).

CF-based product recommendations adopt a user model wherein detailed information is collected about users and their preferences. The data is analyzed in order to determine like-minded users, namely users with similar preferences. The main objective is to create user groups (formally known as clusters) so that the group members share similar preferences. Then, when encountering a user u interested in an item i , where u belongs to a cluster of users with similar preferences, the algorithm may recommend u products liked (i.e. highly rated) by other cluster members. The workflow diagram of the SMARTBUY's CF-based RE is illustrated in Fig. 4.

³ CF is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or 'taste' information from many users (collaborating). The underlying assumption of the CF approach is that, if a person A has the same opinion as a person B on a subject, A is more likely to share B's opinion on a different subject than that of a randomly chosen person.

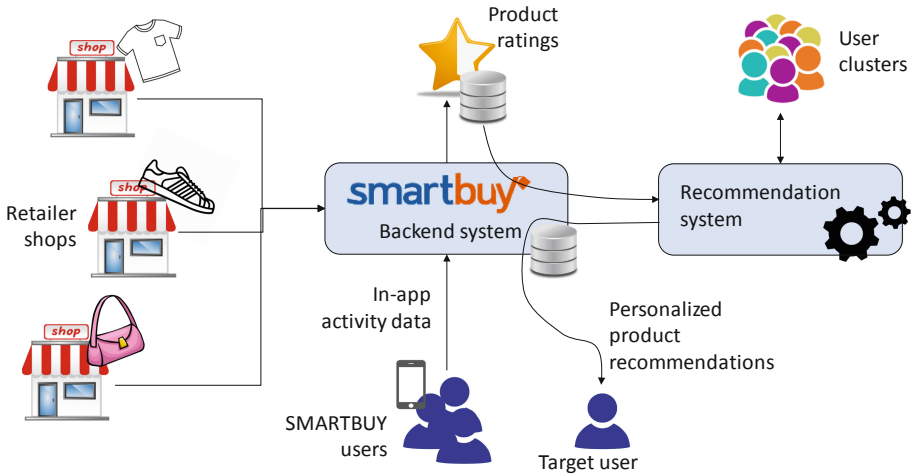


Fig. 4. The SMARTBUY's RE workflow diagram.

It is noted that the user-generated content (in-app activity data) are processed offline in order to form the clusters of users with similar preferences. This is a computationally intensive task, which is typically performed once per day. The output of this computation is the recommendation list of products per user so that the list is instantly delivered to any new (target) mobile user as soon as s/he logs into the system.

5 Pilot Study

The official pilots of SMARTBUY have been executed in 2018, in four European cities with diverse scale and retail market characteristics: London (UK), Patras (Greece), Luleå (Sweden) and Santander (Spain) [23].

An overall number of 471 retailer stores joined SMARTBUY (235 in London, 110 in Patras, 40 in Luleå and 86 in Santander), created an account and uploaded products and promotions in the platform. A total of 2500 individuals (1507 in London, 595 in Patras, 98 in Luleå and 321 in Santander) installed the SMARTBUY app⁴ and acted as customers under real operational conditions. Among them, 91 retailer shop owners and 340 end-users (customers) have accepted to evaluate their overall experience with the platform through responding to an online questionnaire.

Table 1 presents the ten questions (Q) included in the questionnaire handed to 340 participants (customers) of the pilot phase. The table also illustrates a statistical compilation of the responses received (average and median values). It is noted that for questions 1–11 the participants have been requested to reply using the Likert scale (1: Strongly Disagree, 5: Strongly Agree).

Regarding the habits of customers who evaluated the SMARTBUY application, it appears the majority buys products online rather often (Q1). Moreover, customers are

⁴ <https://play.google.com/store/apps/details?id=com.smartbuysshopping.app> .

somewhat divided as regards their preference on the means of delivery for products they have bought online between those that prefer to have them shipped to home and others that prefer to pick them up from the store to save waiting time (and cost) due to shipping (Q2). The preference to collect the purchased products from the store is mostly demonstrated in Patras and Santander, where the customers also tend to prefer buy products from local stores, even if they also practice online shopping (Q3). This finding is also verified by responses to Q4 and Q5 where customers in Patras and Santander expressed their appreciation for being able to collect a product bought online by the physical store, which also results in shipping cost savings. This aspect of SMARTBUY was also appreciated by customers in Luleå, yet, not in London where the longer travel times to reach the physical stores are deterring.

Responses to Q6 and Q7 reveal that the usability of the SMARTBUY application has been found acceptable in all cities. Q8 and Q9 also convey some interesting results. Many users in London, Patras and Santander argued that they discovered new stores, products, sales and services when using the SMARTBUY application, which they would have otherwise missed, revealing its usefulness in online promotion. That was not the case, though, in the smaller-scale city of Luleå, where customers have been rather familiar with most stores.

The last two questions (Q10 and Q11) express the appreciation of customers as regards the performance of the recommendation engine; that is, the accuracy and relevance of recommended products both in terms of the actual type of products and the location of the stores where the products are sold.

Table 1. SMARTBUY evaluation from end-users (customers).

Q#	Questions (customers)	Avg	Median
Q1	I buy products online quite often	3.7	4
Q2	I prefer to collect the products I buy online from the physical store instead of having them shipped to me	3.0	3
Q3	When I shop online, I prefer to buy products from stores that are located in my area	3.4	4
Q4	Booking a product online and receiving it from the physical store is useful	3.4	4
Q5	Booking a product online and receiving it from the physical store helps me save shipping costs	3.4	4
Q6	The SMARTBUY application was fairly easy to use	3.6	4
Q7	The navigation within the SMARTBUY application was easy and comprehensible	3.6	4
Q8	By using the SMARTBUY application I discovered stores which I would not have discovered otherwise	3.7	4
Q9	By using the SMARTBUY application I discovered products, sales and services which I would not have discovered otherwise	3.9	4
Q10	Product recommendations have been relevant to my consumer profile and likings	4.1	4
Q11	Recommended products have been typically offered in stores conveniently located to city areas which I frequently visit	4.2	4

It is noted that the SMARTBUY ecosystem also included a mobile (evaluation) application for assessing the quality of the retailer services [1]. Customers (i.e. pilot users) have had the chance to rate the behavior of retailer store owners as well as the overall provided service. This service has been integrated in the SMARTBUY ecosystem offering an Experimentation-as-a-Service (EaaS) framework [24] developed upon the OrganiCity Experimentation Platform [12]. The evaluation application, on the one hand, is deemed useful to customers to identify retailers with best offers and services, and on the other hand, incentivizes retailers to improve their services.

6 Conclusions

This article presents an overview of the research project SMARTBUY, emphasizing on the motivation behind the research pursued as well as its main outcomes and innovative aspects. SMARTBUY converts physical stores of smart cities in an open, geographically distributed mall by providing the logical consistency needed for conducting centralized searches over independent and geographically distributed physical stores.

SMARTBUY is, firstly, innovative on a conceptual level as it is the first known large-scale experimental system which implements a blended shopping paradigm: it combines the benefits of online shopping with the appeal of traditional store shopping. Moreover, it suggests a roadmap that achieves economies of scale for the, otherwise isolated, small retailers in urban environments, while also offering a suite of innovative services to customers. The core of the system is a module which enables customers to perform location-based searches for products and services offered by member-retailer stores, over a unified, centralized inventory of products.

The project delivers several crucial benefits for retailers:

- achieve economies of scale through enabling access to cloud-based inventory management and digital marketing services;
- generate added value through becoming member of a broad market coalition and taking advantage of geo-marketing opportunities;
- enable broad visibility of offered products via a usable mobile app;
- allow the effective targeting of potential customers through geo-marketing and intelligent context-aware recommendation systems, taking advantage of collected user profile information and user-generated data.

The project also delivers important benefits for potential customers:

- allow one-stop search over a large number of competitive shops in a similar fashion as practiced in brick-and-mortar shopping malls or online shops;
- enable feature and price comparisons, thus, achieving gains in shopping expenditures;
- ensure awareness of -otherwise missed- promotional offers and would-like products, considering a multitude of contextual factors (some of which are not relevant to online customers), such as location & location history, time, season, etc.;

- combine the benefits of online shopping with those of traditional ‘physical’ shopping (social interaction, shopping entertainment, offline retailer loyalty, low product performance risk, prevention of shipment cost and delay, etc.).

Acknowledgement. This work has been partly supported by the University of Piraeus Research Center. The research has also been supported by the EU H2020 Programme under grant agreement no. 687960 (SMARTBUY). The research work of D. Gavalas and T. Chatzidimitris has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK-01572).




References

1. Amaxilatis, D., Giannakopoulou, K.: Evaluating retailers in a smart-buying environment using smart city infrastructures. In: 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 284–288 (2018)
2. Chatzigiannakis, I., Mylonas, G., Vitaletti, A.: Urban pervasive applications: challenges, scenarios and case studies. *Comput. Sci. Rev.* **5**(1), 103–118 (2011)
3. Chen, B.W., Ji, W.: Intelligent marketing in smart cities: crowdsourced data for geo-conquesting. *IT Prof.* **18**(4), 18–24 (2016)
4. Gavalas, D., Kenteris, M.: A web-based pervasive recommendation system for mobile tourist guides. *Pers. Ubiquitous Comput.* **15**(7), 759–770 (2011)
5. Gavalas, D., Konstantopoulos, C., Mastakas, K., Pantziou, G.: Mobile recommender systems in tourism. *J. Netw. Comput. Appl.* **39**, 319–333 (2014)
6. Gensler, S., Neslin, S.A., Verhoef, P.C.: The showrooming phenomenon: it’s more than just about price. *J. Interact. Mark.* **38**, 29–43 (2017)
7. Giffinger, R., Fertner, C., Kramar, H., Meijers, E.: City-ranking of European medium-sized cities. *Cent. Reg. Sci. Vienna UT*, 1–12 (2007)
8. Hsiao, M.H.: Shopping mode choice: physical store shopping versus e-shopping. *Transp. Res. Part E: Logistics Transp. Rev.* **45**(1), 86–95 (2009)
9. Kowatsch, T., Maass, W.: In-store consumer behavior: how mobile recommendation agents influence usage intentions, product purchases, and store preferences. *Comput. Hum. Behav.* **26**(4), 697–704 (2010)
10. Vinod Kumar, T.M., Dahiya, B.: Smart economy in smart cities. In: Vinod Kumar, T.M. (ed.) *Smart Economy in Smart Cities*. ACHS, pp. 3–76. Springer, Singapore (2017). https://doi.org/10.1007/978-981-10-1610-3_1
11. Lin, C., Hong, C.: Using customer knowledge in designing electronic catalog. *Expert Syst. Appl.* **34**(1), 119–127 (2008)
12. OrganiCity project. <http://organicity.eu/>
13. Pantano, E., Timmermans, H.: What is smart for retailing? *Procedia Environ. Sci.* **22**, 101–107 (2014)
14. Pantano, E., Priporas, C.V.: The effect of mobile retailing on consumers’ purchasing experiences: a dynamic perspective. *Comput. Hum. Behav.* **61**, 548–555 (2016)
15. Pantano, E., Rese, A., Baier, D.: Enhancing the online decision-making process by using augmented reality: a two country comparison of youth markets. *J. Retail. Consum. Serv.* **38**, 81–95 (2017)

16. Piotrowicz, W., Cuthbertson, R.: Introduction to the special issue information technology in retail: toward omnichannel retailing. *Int. J. Electron. Commer.* **18**(4), 5–16 (2014)
17. Sanyal, P., Ghosh, A.: Attractiveness of retail agglomeration based on product type: an experimental study. Available at SSRN 2989281 (2017)
18. Sapiezynski, P., Stopczynski, A., Gatej, R., Lehmann, S.: Tracking human mobility using WiFi signals. *PLoS ONE* **10**(7), e0130824 (2015)
19. Sassi, I.B., Mellouli, S., Yahia, S.B.: Context-aware recommender systems in mobile environment: on the road of future research. *Inf. Syst.* **72**, 27–61 (2017)
20. SMARTBUY Deliverable 2.8: SMARTBUY system – final prototypes (2019)
21. SMARTBUY Deliverable 3.2: Wireless geo-located marketing tool (2017)
22. SMARTBUY Deliverable 4.4: Integration of advanced tools for products digitalization and monitoring (2017)
23. SMARTBUY Deliverable 5.4: Deliverable 5.4 report on feedback from real-life customers and retailers (2019)
24. Theodoridis, E., Mylonas, G., Chatzigiannakis, I.: Developing an IoT smart city framework. In: *IISA 2013*, pp. 1–6 (2013)
25. Yang, W.S., Cheng, H.C., Dia, J.B.: A location-aware recommender system for mobile shopping environments. *Expert Syst. Appl.* **34**(1), 437–445 (2008)
26. Yuan, S.T., Tsao, Y.W.: A Recommendation mechanism for contextualized mobile advertising. *Expert Syst. Appl.* **24**(4), 399–414 (2003)



Action Recognition Using Local Visual Descriptors and Inertial Data

Taha Alhersh¹(✉) , Samir Brahim Belhaouari² ,
and Heiner Stuckenschmidt¹ 

¹ Data and Web Science Group, University of Mannheim, Mannheim, Germany
{taha,heiner}@informatik.uni-mannheim.de

² College of Science and Engineering, Hamad Bin Khalifa University,
Education City, Doha, Qatar
sbelhaouari@hbku.edu.qa

Abstract. Different body sensors and modalities can be used in human action recognition, either separately or simultaneously. Multi-modal data can be used in recognizing human action. In this work we are using inertial measurement units (IMUs) positioned at left and right hands with first person vision for human action recognition. A novel statistical feature extraction method was proposed based on curvature of the graph of a function and tracking left and right hand positions in space. Local visual descriptors have been used as features for egocentric vision. An intermediate fusion between IMUs and visual sensors has been performed. Despite of using only two IMUs sensors with egocentric vision, our classification result achieved is 99.61% for recognizing nine different actions. Feature extraction step could play a vital step in human action recognition with limited number of sensors, hence, our method might indeed be promising.

Keywords: Human action recognition · IMUs · Visual descriptors · Feature extraction · Classification · Sensor fusing

1 Introduction

Human behavior analysis tasks are classified according to the degree of semantic as follow: motion, action, activity and behavior [28]. From one hand, motion has the lowest degree of semantic while behavior has the highest one. On the other hand, motion requires the shortest period of time to be done, however to develop a behavior, more longer time of motion capturing is needed. Motion information over time produces action, and different inter-actions construct an activity, and more complex activities shape a behavior.

Recognizing human activity can be based on different sensor modalities, the most common ones are; visual and inertial sensing. Those modalities could be used simultaneously or independently. Inertial measurement sensors (IMUs) is a device with capabilities to measure and report body's specific force, angular rate and orientation of the body. In the sensor's local coordinate system there

are three main measurements: *Accelerometers*: which is the instantaneous acceleration for each axis, *Gyroscopes*: represent the rotational velocity of the inertial and *Magnetometer*: exemplify the instantaneous measured magnetic field with corresponding X , Y and Z axes. One of the drawbacks of using IMUs is the high measurement of uncertainty at slow motion and lower relative uncertainty at high velocities. On the other hand, inertial sensors are able to measure very high velocities and accelerations.

Depending on their methodological nature, two main approaches can be applied on egocentric activity recognition: object and motion based approaches [8, 28]. Object-based approach deals with various information about objects and their interaction with hands. A relationship between object, hand, location and pose will be instantiated to recognize an activity. However, motion-based approach depends on camera location on subjects' body, for example head, shoulder or chest mounted camera. One of the motion-based approaches is optical flow, which refers to the displacement of intensity patterns [20], it represent the motion of visual features such as points, objects, shapes *etc.* via continuous view of the environment to produce a relative motion representation between environment and observer [2]. Optical flow is computed from visual sensing, and can be defined as the apparent motion of objects in consecutive frame pairs. Optical flow can be categorized as forward optical flow when displacement vector for each pixel of the first frame has been computed, or backward optical flow when it's estimated from the second frame. A field of vectors will be generated in u and v directions. Research paradigms in optical flow field are divided into two main methods; first, considering optical flow estimation as a classical problem [10] and is considered as a variational optimization problem to find pixel correspondences between any two consecutive frames [21], or can be formulated as machine learning problem, for example, convolutional neural networks (CNN) [4, 18, 37, 46]. Action frames can be modeled as 3D volumes in time, various local visual descriptors can be extracted like: Histogram of Oriented Gradients (HOG), Histogram of Optical Flow (HOF) and Motion Boundary Histograms (MBH) [15, 26, 42, 43].

The recognition of human actions can be improved by combining different sensor modalities [1, 3, 11]. In this paper, we are combining IMUs data from only two sensors placed at left and right hands with egocentric vision for recognizing similar, complex and opposite actions. Many previous research has focused on recognizing actions that are distinct and independent like: lie, sit, walk, cycle, . . . *etc.* [48] or similar action but with different visual items as; pour bag, pour oil, stir big bowl, stir egg, . . . *etc.* [25]. The main objective of this paper is to recognize complex actions that have similar but opposite nature for example (open, close). This research has four main contributions: First, developing an action extractor tool for both visual and IMUs data based on [49] annotations. Second, introduce a novel statistical feature extraction method for IMUs data based on curvature of function graph and tracking the positions of left and right hands in space. Third, testing complex and similar actions for the same object for example (close-drawer, open-drawer). The last contribution is providing an experimental proof of the limitation of IMUs data to distinguish

actions and suggesting that local visual descriptors can be complementary to IMUs for action recognition.

2 Related Work

Many researchers use only IMUs sensors for action recognition [6, 7, 9, 17, 23, 27, 38, 48]. For example Convolutional Long Term Memory (LSTM) was used to solve sequential human activity recognition problem through proposing a multi-level neural network structure model based on the combination of Inception Neural Network and gated recurrent units (GRU) [48]. The best F-measure score achieved was 94.6% on Opportunity dataset. Bevilacqua *et al.* [9] used Convolutional Neural Networks (CNNs) to classify human activities. They have used Otago exercise program dataset, which contains 16 activities based on five sensors placed on subjects: two sensors placed on the distal third of each shank(left and right), two sensors centered on both left and right feet and one sensor placed on the lumbar region.

In their work Jalloul *et al.* [23] constructed a structural connectivity network to explore the relations between the sensing modules while performing activities based on the correlation between some wearable sensing modules positioned at different parts of the body that constitute a monitoring system for four different activities (walking, standing, lying and sitting). LSTM also was used by [6] to classify 7 main activities with different motion primitives recorded using Apple watch. A hybrid deep framework based on LSTM and extreme learning machine (ELM) was proposed by Sun *et al.* [38] to overcome the problem of sequential activity recognition. The proposed framework composed of convolutional layers, LSTM recurrent layers, and ELM classifier, which can automatically learn feature representations and model the temporal dependencies between features. Their framework have been evaluated on Opportunity dataset with 17 different gestures and achieved 91.8% F_1 score all classes including null class and 90.6% without using null class.

Another work conducted by Rueda *et al.* use CNN using multi-channel time-series for activity recognition and evaluated their model on Opportunity, Pamap2 and Order Picking datasets. A data-driven architecture based on an iterative learning framework was proposed by Davila *et al.* [17] to classify human locomotion activities, such as walk, stand, lie and sit extracted from the Opportunity dataset using multi-class SVM classifier. Their framework produced an average accuracy of 74.08% while using only 6.94% of the samples in the input domain for training compared to average accuracy of 81.07% obtained by the supervised method when using 80% of samples for training and the 20% remaining samples for testing.

On the other hand only Visual data was used in action recognition. For instance, RGB-D data have been used in deep learning to recognize human actions [13, 22]. Sudhakaran *et al.* [35] presented a hierarchical feature lightweight aggregation scheme that can be plugged into any deep architecture with CNN backbone. At each layer the feature from a CNN block is gated and its residual is transferred to the adjacent branch. They have evaluated their proposed

technique on Something-v1, EPIC-KITCHENS and HMDB51 datasets. The results obtained show improvement by about 24% on Temporal Segment Network (TSN). To extract more precise object-related features to guide 3D CNN training, Wang *et al.* [45] introduced Baidu-UTS object detection model that consists of two parts: the first one is a 3D CNN branch that takes sampled video clips as input and produces a clip feature. The second part extracts the object-related features from the context frames. EPIC-KITCHENS dataset was used in their research to predict verbs, nouns, and actions from the vocabulary for each video segment. Sudhakaran *et al.* [36] proposed a Long Short-Term Attention (LSTA), a recurrent unit that addresses shortcomings of LSTM when the discriminative information in the input sequence can be spatially localized, moreover, they have deployed LSTA into a two stream architecture with cross-modal fusion and evaluated their method on four datasets: GTEA 61, GTEA 71, EGTEA Gaze+ and EPIC-KITCHENS. Their network was trained for multi-task classification with verb, noun and activity supervision. Activity classifier activations were used to control the bias of verb and noun classifiers. Other information extracted from visual data also can be used for action recognition applications. For instance, optical flow [2, 24, 39, 47], deep learned spatial descriptors [33] and dense trajectories based on motion boundary histograms [44] which is considered to be invariance to appearance of the representation [31].

Many research work suggests combining different sensor modalities to improve human action recognition [1, 3, 11]. However, for real life scenarios, realistic and compromised number of modalities should be used. Lu and Velipasalar [25] used LSTM to classify actions using four IMUs sensors corresponding 36 components with egocentric video from CMU Multimodal Activity (CMU-MMAC) Database [41]. Visual and audio sensors was used by [5] for activity recognition. [1] has presented framework for recognizing proprioceptive activities using IMUs egocentric data. They used cross-domain knowledge transfer with a CNN-LSTM to exploit discriminative characteristics of multimodal feature groups provided by stacked spectrograms from the inertial data.

The advantages of combining local visual descriptors and IMUs data is the complementary characteristics of visual descriptors and inertial sensors. For instance, IMUs data have large measurement uncertainty at slow motion and lower relative uncertainty at high velocities. Inertial sensors can measure very high velocities and accelerations. On the other hand, visual descriptors can track features very accurately invariant to appearance of the representation at low velocities. For high velocity, tracking is less accurate since the resolution must be reduced to obtain a larger tracking window with the same pixel size [29].

This work is similar to Stein and McKenna work [34], they have proposed using histograms of relative tracklets (RETLETS) and 3-axis *Accelerometer* data at 50 Hz of devices attached to objects: knife, mixing spoon, small spoon, peeler, glass, oil bottle, and pepper dispenser, and no IMUs sensors was placed on subjects bodies for action recognition. However, our method is differ in many aspects: First, in our research IMUs sensor positions are placed on subjects' bodies, not on objects. Second, number of sensors we are using are only two sensors

placed on subjects left and right hands. Third, we are tracking hands positions in space. Forth, we are using novel statistical feature extraction for IMUs data and modified visual descriptors method [42, 43]. Lastly, we are considering similar and opposite scenarios while they are using: add pepper, add oil, mix dressing, peel cucumber, cut ingredient, place ingredient into bowl, mix ingredients, serve salad onto plate, dress salad and NULL classes.

3 Proposed Method

Our method is illustrated in Fig. 1. For each action, corresponding IMUs and visual data is processed to produce features, then an intermediate fusion for the produced features is applied by concatenating features together and then fetched to a classifier for recognition. Details for each step is provided in subsequent sections.

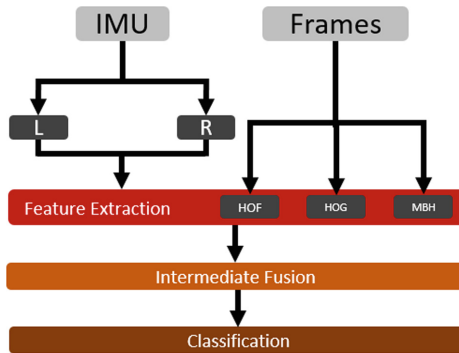


Fig. 1. General overview of our method. For each action, IMUs data are split based on left and right hand sensors, and then extract features. HOF, HOG and MBH for visual data are extracted, then features from IMUs and visual data descriptors are fused to be fetched into classifier.

3.1 Action Extraction

We have developed a Matlab tool¹ to extract the corresponding actions for both IMUs and visual data form CMU Multi-Modal Activity database (CMU-MMAC) [41] based on [49] annotations. The action extraction tool is illustrated in Fig. 2. The user have to provide IDs for IMUs, video, and subject to extract all actions based on the annotation file provided for each subject. Data extracted for both modalities will be synchronized using start and end times provided. To facilitate processing extracted files; each extracted file or image name contains a prefix for subject ID, serial number and action name.

¹ <https://github.com/alhersh/ActionExtractor>.

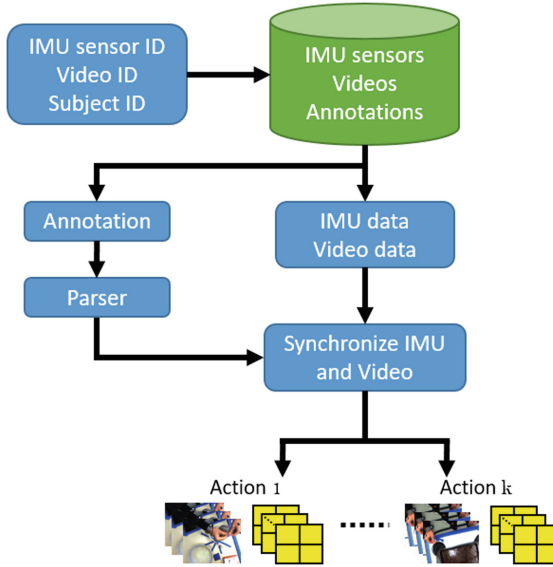


Fig. 2. Action extractor tool. User has to specifying IDs for IMUs sensor, subject and video, to extract the corresponding images and IMUs data for all actions provided based on annotation file.

3.2 Feature Extraction

We have extracted features from both IMUs and visual sensors. From one hand, statistical features for IMUs actions were extracted as shown in Fig. 3, based on a sliding window of size 25 points which is equal to 0.2s and with 40% overlap to produce a feature vector of length 504 as follow:

1. Define left and right sensors.
2. For each left and right hand sensors, the positions of left and right hands is tracked using [40] and the first derivative of the position for each hand is calculated.
3. The difference between the different combinations of left and right hands in 3 dimensional space is calculated.
4. The curvature of (X, Y, Z) acceleration and gyro data is computed using the following formula:

$$k = \frac{|y''|}{(1 + y'^2)^{\frac{3}{2}}} \quad (1)$$

5. Normalize previous values using the formula:

$$f(x) = \frac{x - \mu}{\sigma}, \quad (2)$$

where μ is the mean and σ is the variance.

6. Mean, variance, entropy, kurtosis, moment (order 3 and 4), and number of local maxima (peaks).

We have adopted [42, 43] method for extracting local visual descriptors HOF, HOG and MBH for each action in the video, with two main modifications; the first one is increasing block size from 8 to 50 pixels, according to [14, 15], the recommended values for the HOG parameters are:

- Detection window size is 64×128
- Block size is 16×16

since we are using video resolution of (300×400) , the detection window has been increased by ratio of $468.75\% \times 312.5\%$ and this will increase block size to be 50×50 pixels. The second one, descriptors for each action was aligned with IMUs features to produce a feature vector of length 144 for each action.

Sensors fusion can be performed in three main levels: low level, in which raw data from different sources are combined to produce new more informative data than the inputs. The second level is intermediate level by combining various features from different sensors together to build a feature map. The last level is: high level or can be called decision level, combines decisions from several experts using various methods like; voting, fuzzy-logic and statistical methods [16, 19]. In this work we are using intermediate level of sensor fusion since we are fusing features generated from IMUs and visual data to produce feature vector of length 648 for each action.

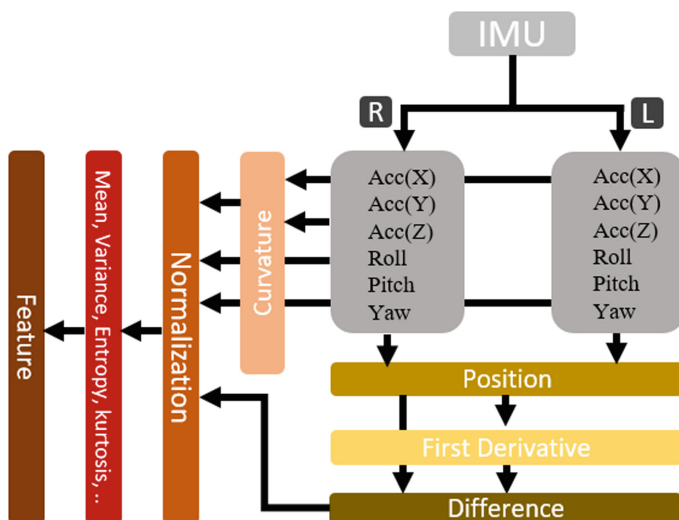


Fig. 3. IMUs feature extraction procedure.

4 Experiments

4.1 Dataset

As has been mentioned earlier in Sect. 2, researchers have used many evaluation datasets for action recognition. In this experiment we have used the CMU Multi-Modal Activity database (CMU-MMAC) [41] since it's the only dataset that combines egocentric vision and IMUs sensors positioned on subjects left and right hands. CMU-MMAC dataset contains measurements of human activities recorded using various modalities: videos including static and wearable cameras, multi-channel audio, IMUs, Radio Frequency Identification (RFID) reader, eWatch, physiological data and motion capturing. More than forty subjects were involved in cooking and food preparation of five different recipes: brownies, pizza, sandwich, salad and scrambled eggs. Based on semantic annotations of CMU-MMAC dataset proposed by [49] three recipes has been annotated: brownie, eggs and sandwich. We have used data extracted from the following modalities in our experiment: First, head mounted high spatial resolution (800×600) camera at low temporal resolution (30 Hz). The resolution of egocentric videos has been reduced by factor of 0.5 to get a resolution of (400×300). Second, two wired IMUs (3DMGX) on right and left hands each of which has a triaxial *Accelerometer*, *Gyro* and *Magnetometer* sensors with sampling rate at 125 Hz, the location of the used sensors in this experiment is plotted in Fig. 4, sample signal is shown in Fig. 5. In this research, we have considered 9 actions with 3 opposite pairs including: (close-bread-bag, open-bread-bag), (close-drawer, open-drawer) and (close-fridge, open-fridge), and fill-oil-oil-bottle-pan, shake-butter-spray-can and stir-bowl-fork.

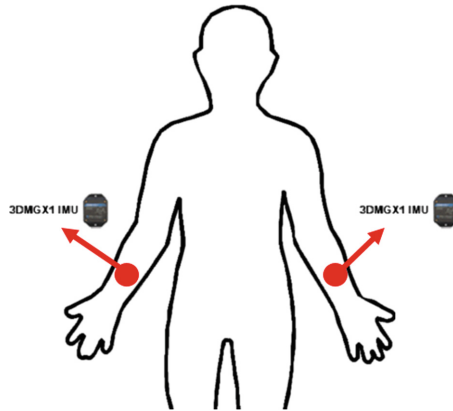


Fig. 4. Location of IMUs sensors (3DMGX) on subjects' right and left hands.

The variability, both between subjects (intersubject variability) and within subjects (intrasubject variability), of the execution time for the same action is

illustrated in Fig. 6. Occurrences on X-axis represents the repetition of action for all subjects performing the same action. While the Y-axis represents the execution time of action in seconds. For instance, action clean can be done in less than 1 s and it can be performed in more than 7 s. Those scenarios increased the complexity of the problem.

4.2 Classification

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification or regression problems. It's a discriminative classifier formally defined by a separating hyperplane. For example, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two dimensional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side. Nevertheless, it's mostly used in classification problems [12].

SVM algorithms use a set of mathematical functions that are defined as kernels. The function of kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions, for example, linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid. In this work we have used Cubic SVM classifier in this research. Based on Matlab 2018b Classification Learner App, the following settings were used for classification process:

- Kernel function: cubic polynomial kernel given the following formula:

$$k(x_1, x_2) = (x_1^T x_2 + 1)^3 \quad (3)$$

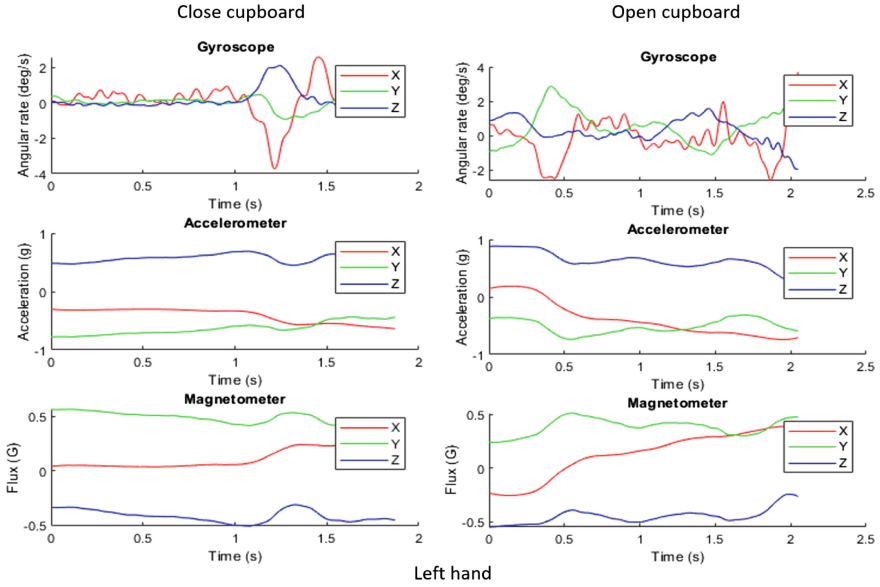
- Kernel scale: automatic
- Box constraint level: 1
- Multiclass method: one-vs-one
- Standardize data: true

Data fetched into the classifiers was split into 80% for training and 20% for testing. We have used 5-fold cross validation for training.

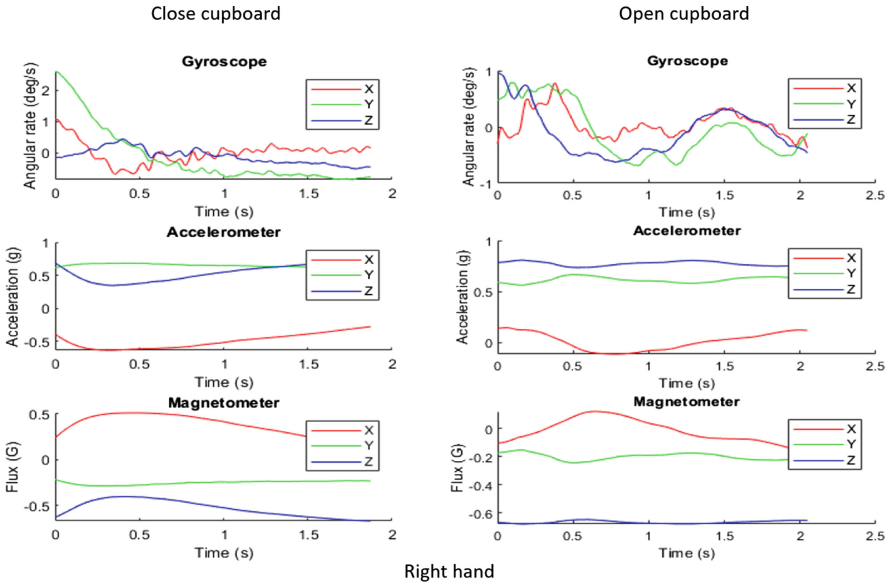
5 Results and Discussion

In this section we are reporting results in Table 1. The first part is reporting classification results for IMUs alone. In the next section we are reporting results for action recognition using visual descriptors only. The third part showing action recognition results using fusing IMUs features and visual descriptors. The last part is a general discussion about the achieved results.

Our recognition result achieved 49.89% for action recognition using only IMUs data from left and right hands. This result is better than [25] by around 5% having in consideration that we are recognizing 9 actions from only 2 sensors while [25] method was used to recognize only 6 action with 4 IMUs sensors, and



(a) Left hand IMU signal



(b) Right hand IMU signal

Fig. 5. Sample signal of two opposite actions for two IMUs sensors: close cupboard and open cupboard for left (a) and right (b) hands for the same subject.

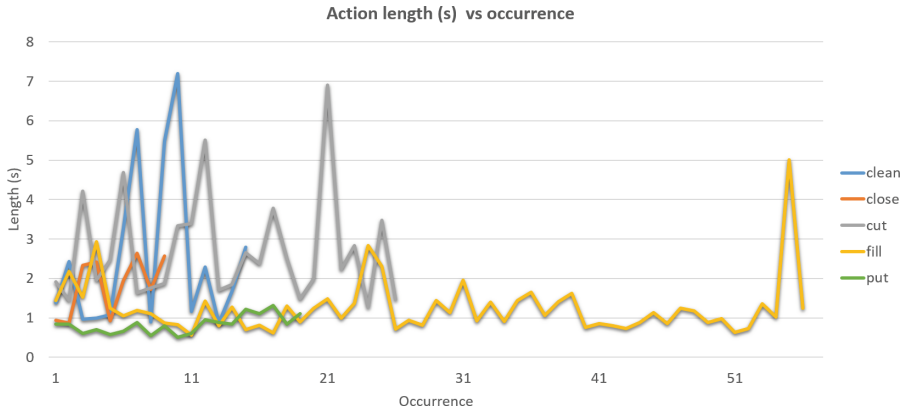


Fig. 6. The variation of execution time for the same performed action. Occurrences on X-axis represents the repetition of action for all subjects performing the same action. While the Y-axis represents the execution time for action in seconds.

they are using LSTM for feature extractions and we are using statistical feature extraction method. On the other hand, [34] achieved 62% f-measure recognition score by using only *Accelerometer* data captured from IMUs sensors attached to non-human objects.

Visual action recognition using VGG16 [32] achieved 58.78% and 64.93% using CapsNet [30]. HOG and HOF were used by [43] for action recognition and HOF has better classification results than HOG. Our visual descriptor results are better than [43] and [34] using HOG and MBH descriptors, but [43] has better results than ours using HOF. The best classification result for recognizing all actions using visual descriptors was 86.88% using our HOG descriptor.

As the literature suggests, combining more than one sensor modality can increase the recognition accuracy. This have been proofed via many previous research. We are working on enhancing the recognition accuracy and our approach was able to achieve this by big margin compared to other methods. IMUs features are fused with the best visual descriptor (HOG) features to produce a feature vector of 648 length for each action. Classifier used in this experiment have been trained to recognize 9 actions, with 3 opposite cases. The results achieved is 99.61% accuracy. The confusion matrix is shown in Fig. 7.

Identifying block size is a major step in local visual descriptors (HOG and HOF) calculation. As [14, 15] suggests that for a detection window of size 64×128 the recommended block size is 16×16 . [43] has used a detection window of size 320×240 and block size of 8×8 . While [34] has used detection window 640×480 with block size 32×32 . In our settings, we have used a detection window of size 300×400 with block size 50×50 and thus our settings are best aligned with block size suggestion by [14, 15] in regard with the detection window size and this explains the differences in action recognition accuracy.

True Class	close-bread_bag	2729								
	close_drawer		982	1						
	close_fridge	3		2319				3		
	fill-oil-oil_bottle-pan	30			2700					
	open-bread_bag	13				4948				
	open-drawer	9					860			
	open-fridge							1334		
	shake-butter_spray_can								562	
	stir-bowl-fork	11								2005
			close-bread_bag	close_drawer	close_fridge	fill-oil-oil_bottle-pan	open-bread_bag	open-drawer	open-fridge	shake-butter_spray_can
		Predicted class								

Fig. 7. Confusion matrix for action recognition results using IMUs and HOG fusion.

Table 1. Comparison of recognition results using IMUs data, visual data, and fusion between IMUs and visual data between our method and three other methods.

Method/dataset	Modalities	Features	Classification accuracy
[25] CMU-MMAC	IMUs	LSTM	45.06
	Visual	VGG16	58.78
	Visual	CapsNet	64.93
	IMUs + Visual	VGG16 + LSTM	65.60
	IMUs + Visual	CapsNet + LSTM	84.40
[43] UCF50	Visual	HOG	76.50
	Vision	HOF	79.50
[34] 50 Salads	Visual	HOG	49.00*
	Visual	HOF	47.00*
	Visual	MBH	53.00*
	IMUs**	Statistics	62.00*
	IMUs** + Visual	(HOG + HOF + MBH) + Statistics	71.00*
Our CMU-MMAC	IMUs	Statistics	49.89
	Visual	HOF	63.29
	Visual	HOG	86.88
	Visual	MBH	71.71
	IMUs + Visual	Statistics + HOG	99.61

*f-measure score.

**IMUs sensors are attached to objects not humans. And, only *Accelerometer* data has been used.

6 Conclusion

In this work, we have developed an action extraction tool for CMU-MMAC dataset based on [49] annotations. Moreover, we have tried to simulate a more realistic scenario that can be used in real life applications using only two IMUs sensors from left and right hands with egocentric vision using opposite actions to make the experiment more realistic. Our results show that intermediate fusion between IMUs and egocentric data improves results by big margin.

One opened question is how to tackle the big temporal variation among the same action occurrences. Also, more investigation is needed when considering opposite actions that have been abandoned by many research works. Future work will focus on using better visual descriptors representations for egocentric data. Moreover, exploring different ways of feature extraction for both IMUs and visual data is needed. Minimizing number of sensors placed on human body for monitoring purpose is needed when considering future smart homes and cities.

References

1. Abebe, G., Cavallaro, A.: Inertial-vision: cross-domain knowledge transfer for wearable sensors. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1392–1400 (2017)
2. Akpinar, S., Alpaslan, F.N.: Video action recognition using an optical flow based representation. In: IPCV, the Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), p. 1 (2014)
3. Alhersh, T., Stuckenschmidt, H.: On the combination of IMU and optical flow for action recognition. In: 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). IEEE (2019)
4. Alhersh, T., Stuckenschmidt, H.: Unsupervised fine-tuning of optical flow for better motion boundary estimation. In: Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Prague, Czech Republic, 25–27 February 2019. VISAPP, vol. 5. pp. 776–783. SciTePress, Setúbal (2019). <https://doi.org/10.5220/0007343707760783>. <http://ub-madoc.bib.uni-mannheim.de/49196/>, online-Resource
5. Arabacı, M.A., Özkan, F., Surer, E., Jančovič, P., Temizel, A.: Multi-modal egocentric activity recognition using audio-visual features. arXiv preprint [arXiv:1807.00612](https://arxiv.org/abs/1807.00612) (2018)
6. Ashry, S., Elbasiony, R., Gomaa, W.: An LSTM-based descriptor for human activities recognition using IMU sensors. In: Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics, ICINCO, vol. 1, pp. 494–501 (2018)
7. Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L., Amirat, Y.: Physical human activity recognition using wearable sensors. *Sensors* **15**(12), 31314–31338 (2015)
8. Betancourt, A., Morerio, P., Regazzoni, C.S., Rauterberg, M.: The evolution of first person vision methods: a survey. *IEEE Trans. Circuits Syst. Video Technol.* **25**(5), 744–760 (2015)

9. Bevilacqua, A., MacDonald, K., Rangarej, A., Widjaya, V., Caulfield, B., Kechadi, T.: Human activity recognition with convolutional neural networks. In: Brefeld, U., et al. (eds.) ECML PKDD 2018. LNCS (LNAI), vol. 11053, pp. 541–552. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-10997-4_33
10. Brox, T., Malik, J.: Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(3), 500–513 (2011)
11. Chen, C., Jafari, R., Kehtarnavaz, N.: A survey of depth and inertial sensor fusion for human action recognition. *Multimed. Tools Appl.* **76**(3), 4405–4425 (2017)
12. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
13. Coskun, H., Tan, D.J., Conjeti, S., Navab, N., Tombari, F.: Human motion analysis with deep metric learning. arXiv preprint [arXiv:1807.11176](https://arxiv.org/abs/1807.11176) (2018)
14. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection (2005)
15. Dalal, N., Triggs, B., Schmid, C.: Human detection using oriented histograms of flow and appearance. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 428–441. Springer, Heidelberg (2006). https://doi.org/10.1007/11744047_33
16. Dasarathy, B.V.: Sensor fusion potential exploitation-innovative architectures and illustrative applications. *Proc. IEEE* **85**(1), 24–38 (1997)
17. Davila, J.C., Cretu, A.M., Zaremba, M.: Wearable sensor data classification for human activity recognition based on an iterative learning framework. *Sensors* **17**(6), 1287 (2017)
18. Dosovitskiy, A., et al.: FlowNet: learning optical flow with convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2758–2766 (2015)
19. Elmenreich, W.: An introduction to sensor fusion. Vienna University of Technology, Austria 502 (2002)
20. Fortun, D., Bouthemy, P., Kervrann, C.: Optical flow modeling and computation: a survey. *Comput. Vis. Image Underst.* **134**, 1–21 (2015)
21. Horn, B.K., Schunck, B.G.: Determining optical flow. *Artif. Intell.* **17**(1–3), 185–203 (1981)
22. Ijjina, E.P., Chalavadi, K.M.: Human action recognition in RGB-D videos using motion sequence information and deep learning. *Pattern Recogn.* **72**, 504–516 (2017)
23. Jalloul, N., Porée, F., Viardot, G., LHostis, P., Carrault, G.: Activity recognition using complex network analysis. *IEEE J. Biomed. Health Inform.* **22**(4), 989–1000 (2018)
24. Kumar, S.S., John, M.: Human activity recognition using optical flow based feature set. In: 2016 IEEE International Carnahan Conference on Security Technology (ICCST), pp. 1–5. IEEE (2016)
25. Lu, Y., Velipasalar, S.: Human activity classification incorporating egocentric video and inertial measurement unit data. In: 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), pp. 429–433. IEEE (2018)
26. Moutinho, N.M.B.: Video and image match searching, US Patent App. 15/252,142, 2 March 2017
27. Moya Rueda, F., Grzeszick, R., Fink, G., Feldhorst, S., ten Hompel, M.: Convolutional neural networks for human activity recognition using body-worn sensors. In: Informatics, vol. 5, p. 26. Multidisciplinary Digital Publishing Institute (2018)
28. Nguyen, T.H.C., Nebel, J.C., Florez-Revuelta, F., et al.: Recognition of activities of daily living with egocentric vision: a review. *Sensors* **16**(1), 72 (2016)

29. Romero, H., Salazar, S., Lozano, R., Benosman, R.: Fusion of optical flow and inertial sensors for four-rotor rotorcraft stabilization. *IFAC Proc. Vol.* **40**(15), 209–214 (2007)
30. Sabour, S., Frosst, N., Hinton, G.E.: Dynamic routing between capsules. In: *Advances in Neural Information Processing Systems*, pp. 3856–3866 (2017)
31. Sevilla-Lara, L., Liao, Y., Guney, F., Jampani, V., Geiger, A., Black, M.J.: On the integration of optical flow and action recognition. arXiv preprint [arXiv:1712.08416](https://arxiv.org/abs/1712.08416) (2017)
32. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
33. Singh, S., Arora, C., Jawahar, C.: First person action recognition using deep learned descriptors. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2620–2628 (2016)
34. Stein, S., McKenna, S.J.: Recognising complex activities with histograms of relative tracklets. *Comput. Vis. Image Underst.* **154**, 82–93 (2017)
35. Sudhakaran, S., Escalera, S., Lanz, O.: Hierarchical feature aggregation networks for video action recognition. arXiv preprint [arXiv:1905.12462](https://arxiv.org/abs/1905.12462) (2019)
36. Sudhakaran, S., Escalera, S., Lanz, O.: LSTA: long short-term attention for ego-centric action recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9954–9963 (2019)
37. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: PWC-NET: CNNs for optical flow using pyramid, warping, and cost volume. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8934–8943 (2018)
38. Sun, J., Fu, Y., Li, S., He, J., Xu, C., Tan, L.: Sequential human activity recognition based on deep convolutional network and extreme learning machine using wearable sensors. *J. Sensors* **2018** (2018)
39. Sun, S., Kuang, Z., Sheng, L., Ouyang, W., Zhang, W.: Optical flow guided feature: a fast and robust motion representation for video action recognition. In: *CVPR* (2018)
40. X-IO Technologies Limited: X-IO technologies limited. UK company (2019). <http://x-io.co.uk/>
41. De la Torre, F., et al.: Guide to the Carnegie Mellon university multimodal activity (CMU-MMAC) database. *Robotics Institute*, p. 135 (2008)
42. Uijlings, J., Duta, I.C., Sangineto, E., Sebe, N.: Video classification with densely extracted HOG/HOF/MBH features: an evaluation of the accuracy/computational efficiency trade-off. *Int. J. Multimed. Inf. Retr.* **4**(1), 33–44 (2015)
43. Uijlings, J.R., Duta, I.C., Rostamzadeh, N., Sebe, N.: Realtime video classification using dense HOF/HOG. In: *Proceedings of International Conference on Multimedia Retrieval*, p. 145. *ACM* (2014)
44. Wang, H., Kläser, A., Schmid, C., Cheng-Lin, L.: Action recognition by dense trajectories. In: *CVPR 2011-IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3169–3176. *IEEE* (2011)
45. Wang, X., Wu, Y., Zhu, L., Yang, Y.: Baidu-UTS submission to the EPIC-kitchens action recognition challenge 2019. arXiv preprint [arXiv:1906.09383](https://arxiv.org/abs/1906.09383) (2019)
46. Wannenwetsch, A.S., Keuper, M., Roth, S.: ProbFlow: joint optical flow and uncertainty estimation. In: *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 1182–1191. *IEEE* (2017)
47. Wrzalik, M., Krechel, D.: Human action recognition using optical flow and convolutional neural networks. In: *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 801–805. *IEEE* (2017)

48. Xu, C., Chai, D., He, J., Zhang, X., Duan, S.: InnoHAR: a deep neural network for complex human activity recognition. *IEEE Access* **7**, 9893–9902 (2019)
49. Yordanova, K., Krüger, F.: Creating and exploring semantic annotation for behaviour analysis. *Sensors* **18**(9), 2778 (2018)



CircuitsMaster: An Online End-User Development Environment for IoT Electronics

Ward Seetsen¹, Irene Mavrommati²(✉), and Vassilis-Javed Khan³

¹ CircuitsMaster Eindhoven, Eindhoven, The Netherlands
wardseetsen@circuitsmaster.com

² Computer Technology Institute and Press “Diophantus”,
Hellenic Open University, Patras, Greece
mavrommati@eap.gr

³ Eindhoven University of Technology, Eindhoven, The Netherlands
v.j.khan@tue.nl

Abstract. Even though Arduino has made creating products with electronics more accessible, a significant number of users still have difficulties with it. An online tool CircuitsMaster.com (CM), aiming make the design of electronics with Arduino faster and easier is presented in this paper. Three diverse needs of designers that wish to include electronics with Arduino in their projects are explained. CM uses a combination of end-user development paradigms to answer to these needs. Based on results from a user evaluation, those who used CM were significantly faster in creating typical electronics assignments when compared to subjects who did not use CM. Therefore, such tools seem to have a salient role to play for designers that wish to develop IoT products.

Keywords: IoT · End-user development · Arduino · CircuitsMaster · Electronics

1 Introduction

Creative people come up with ideas, and incrementally improve on those ideas to an innovative concept that needs to be prototypes. In this ideation process, at a point in time designers need to realize a representation of their concept. This takes the form of a prototype, and is an essential part of the design process of realizing ideas. If the concept involves electronics, the prototype will also need to include electronics. Nevertheless, coming up with a great idea involving electronics, does not necessarily mean the person that came up with that idea has the know-how of electronics. Even if one has the know-how the issue of time and effort spent in realizing the concept comes up. Currently the importance of electronics in the design area is gaining significance, due to the rise of the Internet of Things (IoT) (Barricelli and Valtolina 2015), applications.

In order to make IoT prototypes designers often use Arduino (Fogli et al. 2015). Arduino is an open-source electronics platform that is created to behave as the ‘brain’ of any electronic system.

In Arduino, different circuits with different purposes can be connected, for example a motor, a Bluetooth module etc. This is partly made easier due to companies like

Adafruit and Sparkfun that sell pre-made circuits, containing all the electronic components for a specific function. When it comes to programming Arduino, one can do that, with a variant of a C-based programming language. However, programming an Arduino still requires some specialized programming skills (Barricelli and Valtolina 2015). To make Arduino programming easier, several applications are already created; examples include: Ardublock, miniBloq, S4a, Modkit and Visuino. All these platforms have already made prototyping electronics for designers more accessible, there still a diversity of users' needs that are left unmet, because these platforms can still be experienced as confusing (Barricelli and Valtolina 2015).

This paper contributes to existing literature regarding programming with Arduino, by presenting:

- (1) the diversity of design student needs when it comes to programming electronics based an interview study with nine industrial design students
- (2) CircuitsMaster.com (CM) - a novel online tool for developing electronics, and
- (3) an experiment that shows that industrial design students CM can develop faster electronic circuits with CM when compared to their current practices.

2 Related Work

Prior research has shown that dealing with electronics is a serious bottleneck for designers (Fogli et al. 2015), (Barricelli and Valtolina 2015). A series of interviews done with Industrial Design students were done prior to this study to find these bottlenecks. The main bottlenecks were not knowing how to connect the different components, not knowing how to program for Arduino and not understanding the examples that can be found on the Internet. This appeared to cause frustrations for the students.

Prior research to end user programming for applications for Internet of Things has a history since the early 2000's: Composition of IoT applications can be done interactively, with end-users explicitly involved in the IoT application configuration with the mediation of End User Development (EUD) tools, or automated by intelligent agents (Davidyuk et al. 2015), (Markopoulos et al. 2017). In the latter case (intelligent automation) user involvement is limited, as the system autonomously creates the application's functionality, although there can be limitations regarding the degree to which user expectations are met.

Programming by Example is one strategy used in End User Development (Dey et al. 2004; Chen and Li 2017). The user in in this approach performs specific examples of system behavior and interaction, and the system infers the application logic. Intelligent systems further support generalization and reuse of the same application logic, by inferring patterns in other similar occasions (Chin et al. 2006). Nevertheless, inferring the correct control logic can be a challenge in programming by example strategies, thus pointing at a need for tools targeted to end users, with which that can further detail application parameters and specify rules. To this end, Rule-based programming is often suggested, with (Huang and Cakmak 2015; Ghiani et al. 2017) proposing a Trigger-action programming style, based on event-condition-action rules.

In other approaches, people compose IoT applications interactively, assisted by Tools that facilitate End User Development. Some such tools use metaphors that can map programming constructs to physical/tangible concepts that users know and can apply more easily. Various such metaphors have been recommended in the relevant literature such as join the dots, pipeline, jigsaw puzzle, (Gross and Marquardt 2007) (Danado and Paternò 2015; Davidyuk et al. 2015). (Fogli et al. 2016). (Kameas and Mavrommati 2005) proposed an editing tool promoting a high-level conceptual model - a simple visual representation of a “join-the-dots” metaphor - (Mavrommati et al. 2004), while further rule-based editing enables detailed programming of the parameters of the application.

Smart environments created by putting together off-the-shelf smart devices is discussed in (Kubitza and Schmidt 2015). Users can create new applications by using existing devices, concentrating more on the implementation of the application logic. Application’s behavior can be changed with tools that add new context aware rules. Such as a web-based tool has been designed for managing the different application components, and expresses the application logic, such as conditions and events, with JavaScript. This tool seems therefore more fit to be used by programmers rather to other, less computer savvy end-users.

The conventional way for a user to instruct Arduino, is a programming language. Yet programming languages are primarily targeting software engineers and not designers. The field of End-User Development wishes to diversify programming to other user groups than professionals. EUD is defined as “A set of methods, techniques, and tools that allow users of software systems, who are acting as non-professional software developers, at some point to create, modify or extend a software artefact” (Barricelli and Valtolina 2015). According to this definition, CM can be defined as an end-user development tool in the area of electronics where designers are the end-users. The use of end-user development where designers are the end-users is not a novel idea. Several applications have already tried to achieve this goal in other areas of expertise and succeeded. For example, WordPress has allowed people other than web-developers to develop websites. This is widely used by designers (About WordPress 2010). Other online platforms like Appsbuilder (apps-builder.com) have also achieved this goal for the development of mobile apps.

Several well-researched programming paradigms of end-user development already exist, as mentioned earlier in this section. Programming by demonstration; visual programming; programming by instruction; and programming by example are the four categories mentioned by (Barricelli and Valtolina 2015).

In programming by demonstration, the user demonstrates how the software should behave. The problem with this paradigm is that many conditions can easily be forgotten (Schmidt 2015), especially for complex applications. This causes programs not to behave as desired. Visual programming is used in most of the currently available end-user development Arduino applications. The user is still programming, but with blocks instead of text. This minimizes the errors, but still works from an engineer’s point of view and not from a designer’s point of view, since the visual blocks are a mere different representation of the programming commands. In programming by instruction, the user inputs rules, instead of programming commands, which describe the desired behavior. An example of this paradigm, is the “IF-THIS-THEN-THAT” (Barricelli and

Valtolina 2015) (ifttt website). It is known as easy to learn (Lucci and Paterno 2015), making this paradigm suitable for CM. In programming by example, the user chooses an example, which closely resembles the desired behavior, and adjusts this example to achieve the exact desired behavior (Lucci and Paterno 2015). Programming by example is also a known way to create complex applications with little required knowledge, making this also applicable for CM. Wherever programming by example is used, it is important to structure the possible options into intuitive logical categories for users (Lucci and Paterno 2015). So, although several well-known paradigms are used in different domains, when it comes to programming IoT products, Arduino is still the state-of-the-art. This research has set off to investigate whether current designers have unmet needs with Arduino and how can we apply well known EUD methods to diversify design of electronic circuits.

3 Assessment

3.1 Preliminary Study: Interview

Before developing CM, nine industrial students were interviewed: five females and four males, ages between 18 and 25, from two different universities (eight from one department, one from another). Interview questions included, among others: current practices and tools they use to develop electronics; how often they create electronics; how long and what information it takes to create them; their frustrations with current tools.

3.2 Evaluation Experiment

Sixteen industrial design students were recruited - 14 from one University and two from another. Their ages ranged from 18 to 26. Their self-reported level of expertise in electronics varied from novice to skilled. One participant was a master student and the others were bachelors, divided over all three bachelor years.

The two research questions were:

- (a) Are designers' success rate in designing electronics higher with CM than without it?
- (b) Can designers create electronic circuits faster with CM than without it?

Although we recruited two different groups of participants all of them follow the same educational program - i.e. industrial design. Furthermore, we chose different groups for practical purposes (i.e. to speed up recruitment) as well as avoid potential biases of the first group we interviewed.

3.3 Application and Workflow Used

The version of CM that they used for the experiment is further described in this section. CM is unique in that based on IF-THEN rules it automatically generates:

- the Arduino circuit in a graphical representation (Fig. 2),

- the list of components that the user needs (Fig. 3), and
- the necessary Arduino code,
- Furthermore, based on the list of components, CM recommends direct links to e-shops that one could purchase those components.

The CM workflow is as follows: the user starts by adding ‘objects’: these are electronic input and output devices i.e. sensors and actuators. For every chosen ‘object’ an example circuit is added to the schematic. This way the user ends up with a full schematic for all the required parts. The circuit is build around an Arduino. The user can add *what* the ‘objects’ should do and *when* these objects should do this i.e. actions and conditions for each action. By setting the actions and conditions for each object, the user gives instructions to CM in an IF-THIS-THAN-THAT way combined with answering the questions CM asks to understand in more detail what the user wants the electronics to do (programming-by-instruction). CM interprets this input and translates it to a combination of code examples to generate a full Arduino code for the user (programming-by-example).

To illustrate the workflow, the following example is used: If the user wants to change an LED color based on the environment temperature, the user first adds an LED. Then the user adds two actions to this LED: turn red and turn blue. Then the user adds a condition to both actions. To the ‘turn red’ action the user adds the condition ‘if the temperature is higher than 20°’ and to the ‘turn blue’ action the user adds the condition ‘if the temperature is lower than 20°’ (Fig. 1). Then the user clicks ‘Generate Circuit’ and (s)he receives a circuit with an LED, a temperature sensor and an Arduino code to change the LED color depending on the temperature.

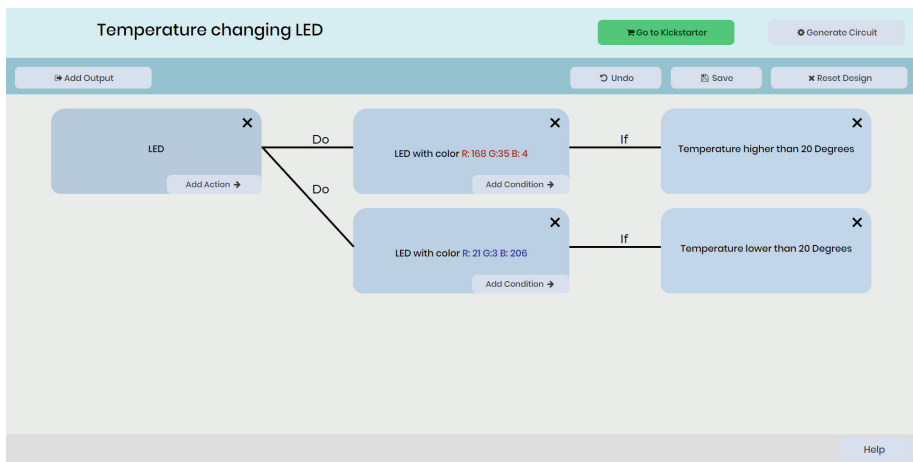


Fig. 1. Circuit design CircuitsMaster

CM is currently available at CircuitsMaster.com. The current version includes 21 components and the premium version includes 50 components. The demo-version used in this experiment is available at <http://circuitsmaster.com/Demo/>.

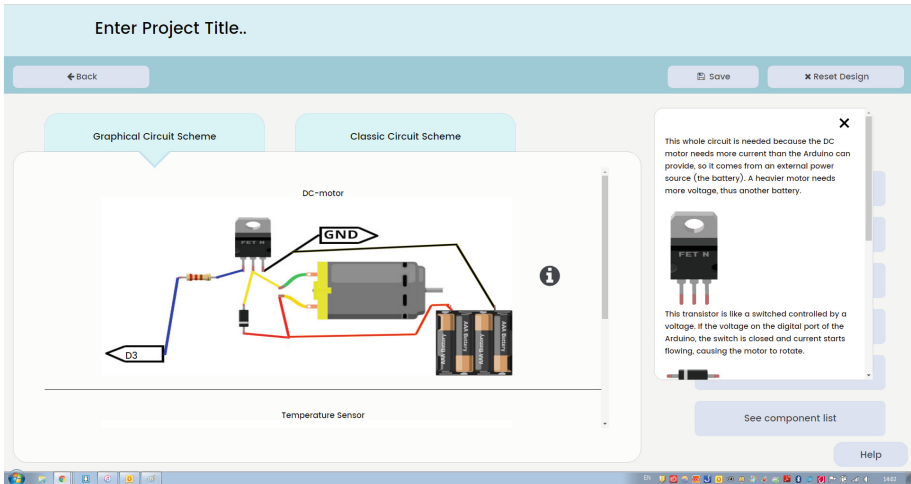


Fig. 2. Circuit generation CircuitsMaster

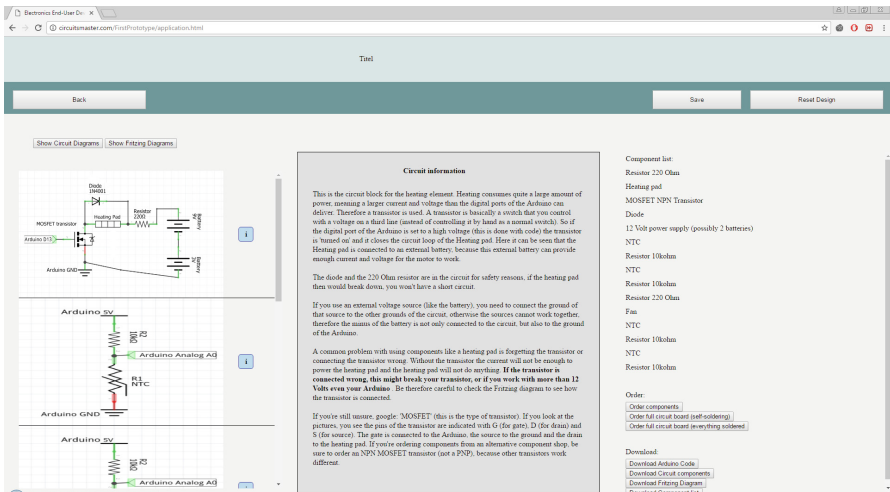


Fig. 3. Screenshot from CircuitsMaster: easy assignment, output

3.4 Evaluation Process

Initially, participants’ skill in electronics were assessed, by asking subjects to grade their skill level and describe about what they had achieved with electronics so far. Based on their answers a participant was classified in the following three categories: novice, intermediate and advanced. Special attention was taken so that in the experiment there was no participation of any experts since they are not CM’s target group.

Depending on their skills category, a participant was given an assignment in their level (i.e. easy, intermediate and complex). The easy assignment was to make a

thermostat. The intermediate to create an environment system that measured temperature, CO gas and methane gas and indicates if the values are safe. The complex assignment was to create a wake up alarm clock that moves away from the user once it goes off. The difficulty was based on the number of components required and the number of connections between the different components. Participants were asked to complete the assignment up to the point that one could start building it. This in effect means having the circuit scheme, the code and knowing which components to use. A between-subjects research design was opted for - i.e. one group of participants used CM and the other was asked to use whatever method they were used to. Ten participants used CM (3 had novice skills, 4 intermediate, 3 advanced). Six did not use CM and relied to whatever tools they were used to (3 intermediate, 3 advanced). Each participant had 20 min to complete the assignment. If the assignment was not completed, participants were asked to estimate the time it would take to finish it. The participant's relative success (in percentage) was also noted when finished. If the participant perfectly completed the assignment, it was assigned a 100%. Otherwise the amount of correct outputs, actions of outputs and inputs were counted and divided by the total amount of outputs, actions and inputs of the correct results. Incorrect actions, inputs or outputs were subtracted from the correct ones. This calculated value then represented the assignment's success percentage for a certain participant. The same method was used for both groups, participants who used CM and those who did not (Fig. 4).

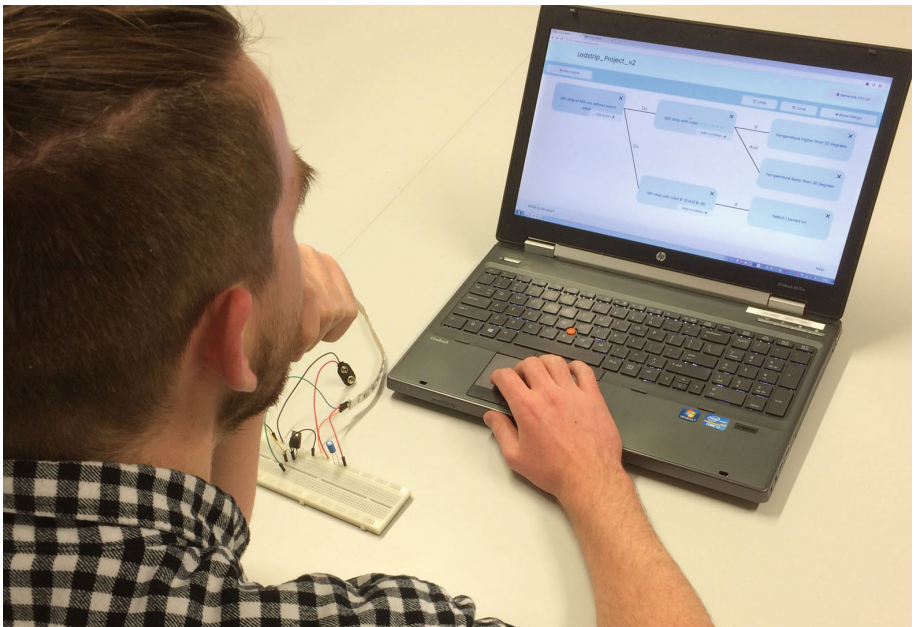


Fig. 4. User using CircuitsMaster

4 Results

4.1 Interviews: Three Diverse Unmet Needs

All test participants agreed on the importance of designers being able to design electronics. They need to know what is available regarding electronics and how to use them, because they currently need them in almost every project. Some are really good in using electronics and some are not, there is a considerable variation in this respect. Similarly, there is a variation in appreciating working with electronics or not; some designers like designing the electronics and some do not. However, the participants agreed that knowledge of electronics should not limit their creativity in design. All participating subjects use Arduino in their student projects and build the rest of the electronics around this platform. The main problems they experience is *programming* the Arduino, but also making the *circuit scheme*. An interesting finding is that making the circuit scheme is generally perceived as harder than programming. Another problem that is less mentioned is *soldering*. It is particularly frustrating when the electronics break while soldering. Furthermore, getting all *electronics compact together* can also be a problem for designers. The estimated time it takes to make the electronics also varies from 15% to 30% of the total project time. After presenting the common findings across all participants, three diverse needs were identified, that are currently unmet for industrial design students. These needs are described by referring to them as “personas”.

Persona 1: Uncertain and Afraid

This persona does not feel confident in making electronics and is afraid to break components. Understanding electronics and different components is a problem and also making a circuit scheme and programming the code. Because of this the use of electronics is generally avoided. The use of electronics in prototypes is experienced as frustrating. CM supports this persona by offering a short explanation for all the circuits and offering the guarantee that every circuit in CM is tested and will work.

Persona 2: Lack of Info

This persona has more knowledge of electronics, but not enough to create the electronics wanted in prototypes. Everything is created from examples that are adjusted afterwards. The motivation to learn in electronics is there, but there is no time for this, the electronics need to be used now. If necessary, the concept is slightly changed and compromised on the spot so that this persona will be able to implement it. This persona is supported by CM, because all the information this persona needs is in one place: the CM-environment.

Persona 3: Get Results Faster

This persona can already make everything she wants. Also examples are used as the basis of all electronics designs and these examples are adjusted to her personal needs. Values of components are not calculated but she does a calculated guess and afterwards measurements to check if the behavior is as desired (the same holds for sensors and actuators). If necessary, the concept is slightly changed to implement it. Although she can get everything working she gets frustrated if it takes too long for actually creating the product. This persona is supported by CM by all the component-values that are already calculated and by pre-made code that CM provides.

4.2 Experiment: CM Speeds up Development

Due to the fact that the Levene's test for equality of variances was statistically significant for both the success rate and the completion time our data's normal distribution cannot be assumed and therefore a Mann-Whitney test was conducted instead of a t-test. The Mann-Whitney test indicated that the success rate of an assignment is higher with CM ($M = 93.10\%$, $SD = 10.77\%$, Mean Rank = 10.10) than without it ($M = 55.83\%$, $SD = 40.58\%$, Mean Rank = 5.83), but it is not statistically significant $U = 14$, $p = .06$. The Mann-Whitney test indicated that the time needed for an assignment with CircuitsMaster ($M = 6.9$ min, $SD = 2.37$, Mean Rank = 5.5) is lower than the time needed for an assignment without CircuitsMaster ($M = 189.17$ min, $SD = 150.94$, Mean Rank = 13.5) and this case the difference is statistically significant $U = 0$, $p < .0001$ -that is including the extra time that participants had estimated beyond the 20 min that we had set for the assignment. The same result is showing if we do not take into account the extra time since all of the 6 participants that did not use CM needed more than 20 min to complete the assignment that they were given.

It was also checked whether the three novice users, who were in the CM group, had an effect in the results. The answer to that check is negative. When filtering out those participants the results are the same. Again success rate (Fig. 5) was higher for the CM group ($M = 95\%$, $SD = 11.18\%$, Mean Rank = 8.71) when compared to the other group ($M = 55.83\%$, $SD = 40.58\%$, Mean Rank = 5) but not statistically significant $U = 14$, $p = .06$. Again completion time was faster with CM ($M = 6.57$ min, $SD = 1.61$, Mean Rank = 4) when compared to the other group ($M = 189.17$ min, $SD = 150.94$, Mean Rank = 10.5) and was statistically significant $U = 0$, $p = .001$.

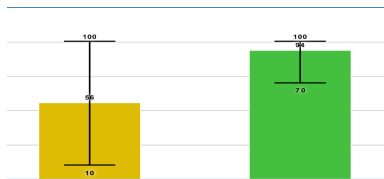


Fig. 5. Success rate with and without CircuitsMaster

Apart from this quantitative data, qualitative data were collected with a debriefing interview. The majority of our participants reported to have enjoyed using the application and stated that it could help them with electronics design.

Some participants also indicated that CM could stimulate their creativity, because it allows them to browse through the possibilities. Participants also stated that the graphical circuit diagrams, made with the Fritzing application (Kraúnig 2009) were preferred over the classic circuit diagrams. Furthermore, when looking at the actual components that our participants used, we observed that participants using CM were more likely to choose components that can be experienced as complex (Fig. 6).

For example, all participants that used CM included a LED strip or neopixel ring (Fig. 7), while all participants who did not use CircuitsMaster included single LEDs. An improvement point that was mentioned was about the graphics of CM – the need

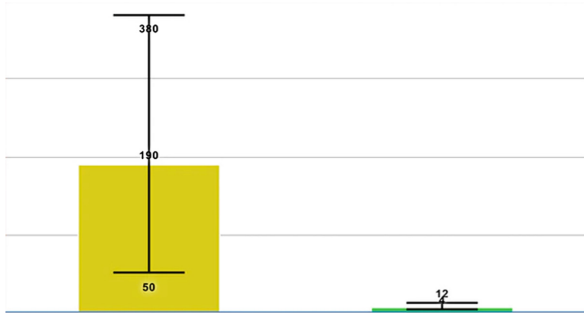


Fig. 6. Time in minutes required for assignment with and without CircuitsMaster

was pointed for the User Interface design to become more professional. Examples include text fields being selected once they appear, shortcuts and being able to drag-and-drop content.

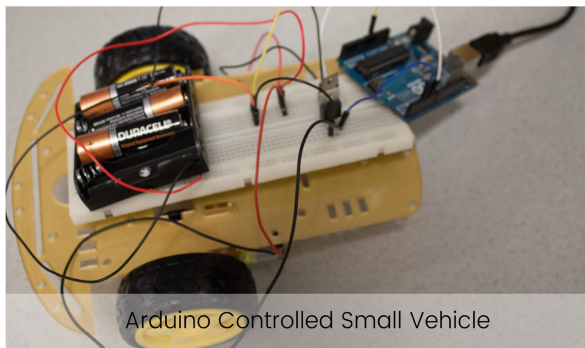
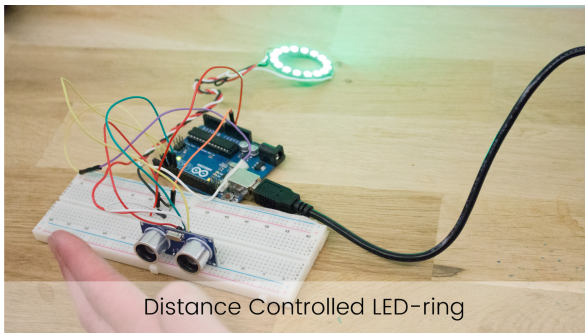


Fig. 7. Resemblance of projects made during the final user test

The participants had divergent opinions about CM’s workflow. Some participants preferred to start with the conditions and work to the output and some participants

preferred it the other way around. Also some participants indicated that they would prefer a drag- and-drop interface. Even though participants had divergent opinions on this topic, they all got used to the application quickly.

5 Discussion - Future Work

The goal of research presented here was to build a tool to help an application to help creative individuals in configuring the electronics in their prototypes. The user tests have shown that CircuitsMaster achieves this in multiple ways, both in succeeding percentage as well as in the amount of time necessary to do these projects. CircuitsMaster is a webtool that has the potential to be added to the toolkit of many designers to help them to make prototypes better and faster. CM as a tool is therefore seen as being a valuable addition to the available tools for electronics design for non-experts. The development of CircuitsMaster will continue since many people can have benefit from this application and a kickstarter campaign has been launched.

There have been limitations of the performed studies: More subjects would be needed so as to draw more valid significant conclusions. One additional limitation is that the first study has been performed with an early prototype of CircuitsMaster. This was done, because the qualitative results of the study were used to improve the prototype and no other prototype was available at the time of the study. It is assumed that the results would be different (improved) with the updated later version, but this assumption has not yet been proven by tests. Repeating the studies with the final version of CircuitsMaster would be a benefit.

The demographics of the participants have been limited. Only results from Industrial Design students from the Technical University of Eindhoven are taken into account in the analysis. This is the case, because the initial target group was Industrial Designers and CircuitsMaster was founded in an Industrial Design Department of the Eindhoven University of Technology. Since the target group Extends further to these profiles, it would be beneficial to redo the full research with the full target group in various locations. Including the full target group means also including participants completely novice in creating electronics, including hobbyist.

From the user studies, several conclusions regarding CircuitsMaster can be drawn: The first one is that users are likely to have a higher success rate when using CircuitsMaster compared to not using CircuitsMaster. Also users can realize their electronics projects faster with CircuitsMaster. Finally, CircuitsMaster is easy to use. CircuitsMaster has also shown that it is possible to generate electrical systems from an input method that is not directly related to the programming language (Fig. 8).

The main author continues working with CircuitsMaster to bring it to a point where it can be used and can live up to the expectations of the target group and to the potential of the application. With the help of the crowdfunding campaign, CircuitsMaster aims to be further realized. Once the application is finished, the next phase will be to improve it and explore the use in the several target groups. In the course of this research it was in its preliminary phase and tested only with Industrial Designers, yet, in later stages the target group can be extended. Testing with different target groups will be able happen after the finalization of the first version to be launched.

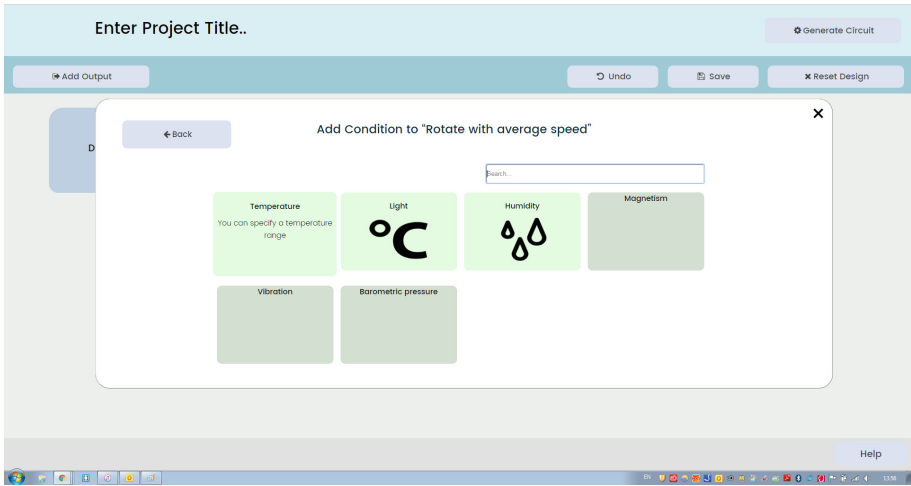


Fig. 8. Instructing CircuitsMaster

6 Conclusions

While recognizing the limited number of participants in this study, two important contributions need to be stressed, that pave the way for future research studies. The first contribution considers the diversity of unmet needs that industrial designers have with regards to designing products with electronics. Based on an interview study ($N = 9$) we identify three diverse profiles of designers that future electronics tools need to support designers in:

- (1) boosting the confidence, especially of novice ones, by clarifying what components are needed and how to connect them;
- (2) improving their knowledge by integrating examples they can consult relating to the specific project they are currently working one;
- (3) making the creation of the circuit faster.

The second contribution considers CircuitsMaster.com (CM), an online tool that is designed to support the three aforementioned needs. In an in-between subjects' experiment with industrial design students ($N = 16$) we showed that CM speeds up considerably the completion time of typical assignments, with different levels of difficulties, that involve electronics.

Furthermore, although not statistically significant, participants who used CM had a higher success rate ($M = 93.10\%$, $SD = 10.77\%$) within a 20 min' window when compared to students that did not use CM ($M = 55.83\%$, $SD = 40.58\%$). These results make us confident that with further improvement CM can become a salient weapon in the arsenal of industrial designers that wish to develop IoT products. Future work includes developing further CircuitsMaster.com. Apart from adding more components, CM will support sharing projects among users, and active guidance in writing code and in actually building the circuit.

References

- Fogli, D., Giaccardi, E., Acerbis, A., Filisetti, F.: Physical prototyping of social products through end-user development. In: Díaz, P., Pipek, V., Ardito, C., Jensen, C., Aedo, I., Boden, A. (eds.) IS-EUD 2015. LNCS, vol. 9083, pp. 217–222. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-18425-8_19
- Barricelli, B.R., Valtolina, S.: Designing for end-user development in the Internet of Things. In: Díaz, P., Pipek, V., Ardito, C., Jensen, C., Aedo, I., Boden, A. (eds.) IS-EUD 2015. LNCS, vol. 9083, pp. 9–24. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-18425-8_2
- Adafruit. <https://www.adafruit.com/>. Accessed July 2019
- Sparkfun. <https://www.sparkfun.com/>. Accessed July 2019
- Davidyuk, O., Milara, I.S., Gilman, E., Rieki, J.: An overview of interactive application composition approaches. *Open Comput. Sci.* **5**(1), 79–95 (2015)
- Markopoulos, P., Nichols, J., Paternò, F., Pipek, V.: End-user development for the Internet of Things. *ACM Trans. Comput.-Hum. Interact. (TOCHI)* **24**(2), 9 (2017)
- Dey, A.K., Hamid, R., Beckmann, C., Li, I., Hsu, D.: a CAPpella: programming by demonstration of context-aware applications. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 33–40. ACM (2004)
- Chen, X.A., Li, Y.: Improv: an input framework for improvising cross-device interaction by demonstration. *ACM Trans. Comput.-Hum. Interact. (TOCHI)* **24**(2), 15 (2017)
- Chin, J.S.Y., Callaghan, V., Clarke, G.: An end-user programming paradigm for pervasive computing applications. In: ICPS, vol. 6, pp. 325–328 (2006)
- Huang, J., Cakmak, M.: Supporting mental model accuracy in trigger-action programming. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 215–225 (2015)
- Ghiani, G., Manca, M., Paternò, F., Santoro, C.: Personalization of context-dependent applications through trigger-action rules. *ACM Trans. Comput.-Hum. Interact. (TOCHI)* **24**(2), 14 (2017)
- Gross, T., Marquardt, N.: CollaborationBus: an editor for the easy configuration of ubiquitous computing environments. In: 15th EUROMICRO International Conference on Parallel, Distributed and Network-Based Processing, PDP 2007, pp. 307–314. IEEE (2007)
- Danado, J., Paternò, F.: A mobile end-user development environment for IoT applications exploiting the puzzle metaphor. *ERCIM News* **101**, 26–27 (2015)
- Fogli, D., Lanzilotti, R., Piccinno, A.: End-user development tools for the smart home: a systematic literature review. In: Streitz, N., Markopoulos, P. (eds.) DAPI 2016. LNCS, vol. 9749, pp. 69–79. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39862-4_7
- Kameas, A., Mavrommati, I.: Extrovert gadgets. *Commun. ACM* **48**(3), 69 (2005)
- Mavrommati, I., Kameas, A., Markopoulos, P.: An editing tool that manages device associations in an in-home environment. *Pers. Ubiquit. Comput.* **8**(3–4), 255–263 (2004)
- Kubitza, T., Schmidt, A.: Towards a toolkit for the rapid creation of smart environments. In: Díaz, P., Pipek, V., Ardito, C., Jensen, C., Aedo, I., Boden, A. (eds.) IS-EUD 2015. LNCS, vol. 9083, pp. 230–235. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-18425-8_21
- About WordPress: Beginning WordPress 3, pp. 1–14 (2010). https://doi.org/10.1007/978-1-4302-2896-7_1
- apps-builder.com: Create & publish powerful native apps (n.d.). <http://www.apps-builder.com/>. Accessed 06 Feb 2017

- Schmidt, A.: Programming ubiquitous computing environments. In: Díaz, P., Pipek, V., Ardito, C., Jensen, C., Aedo, I., Boden, A. (eds.) IS-EUD 2015. LNCS, vol. 9083, pp. 3–6. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-18425-8_1
- Lucci, G., Paternò, F.: Analysing how users prefer to model contextual event-action behaviours in their smartphones. In: Díaz, P., Pipek, V., Ardito, C., Jensen, C., Aedo, I., Boden, A. (eds.) IS-EUD 2015. LNCS, vol. 9083, pp. 186–191. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-18425-8_14
- Kraúnig, A.: Fritzing: a tool for advancing electronic prototyping for designers (2009)
- ifttt.com: If This Then That: a free web-based service to create chains of simple conditional statements, called applets, August 2019



Enabling Machine Learning Across Heterogeneous Sensor Networks with Graph Autoencoders

Johan Medrano¹  and Fuchun Joseph Lin²

¹ LIRMM, Université de Montpellier, Montpellier, France
johan.medrano@laposte.net

² Department of Computer Science, National Chiao Tung University,
Hsinchu, Taiwan

Abstract. Machine Learning (ML) has been applied to enable many life-assisting applications, such as abnormality detection in daily routines and automatic emergency request for the solitary elderly. However, in most cases ML algorithms depend on the layout of the target Internet of Things (IoT) sensor network. Hence, to deploy an application across Heterogeneous Sensor Networks (HSNs), i.e. sensor networks with different sensors type or layouts, it is required to repeat the process of data collection and ML algorithm training. In this paper, we introduce a novel framework leveraging deep learning for graphs to enable using the same activity recognition system across HSNs deployed in different smart homes. Using our framework, we were able to transfer activity classifiers trained with activity labels on a source HSN to a target HSN, reaching about 75% of the baseline accuracy on the target HSN without using target activity labels. Moreover, our model can quickly adapt to unseen sensor layouts, which makes it highly suitable for the gradual deployment of real-world ML-based applications. In addition, we show that our framework is resilient to suboptimal graph representations of HSNs.

Keywords: Graph autoencoders · Heterogeneous sensor networks · Smart homes

1 Introduction

The development of networking technologies and the advance in embedded computing enable the widespread deployment of IoT sensor networks. In this context, pervasive sensing and actuation with ubiquitous Internet of Things (IoT) tends to be a natural direction. The development of ML algorithms with pervasive sensing enables applications that can significantly enhance our daily lives. Some outstanding examples include abnormality detection in the routines of solitary elderly persons [1], assisted living for people with dementia [2, 3], early detection of Parkinson disease [4].

To perform these tasks, ML algorithms can take advantage of a large amount of data available for training. However, in most cases these algorithms are strongly tied to the physical layout of the sensors [5]. This makes their general applicability on real-world applications doubtful. First, due to the dependence on the structure of the sensor

network, the deployment of applications to other sensor networks is impaired as it requires the collection of new data and the training of a new ML model. Consequently, each deployment requires a repetition of the same effort. Second, when the model is trained online with the collected data, the application requires a significant amount of time before achieving decent performances. To reduce the deployment overhead and enable large-scale applications, there is an urgent need for a solution allowing the use of the same ML model across HSNs.

The existing frameworks rely on complex methods often tied with the adapted ML model. With the objective of proposing a simpler approach to adapt ML models, in particular classifiers, across HSNs, we introduce a novel framework in this paper. In the proposed architecture as depicted in Fig. 1, a first component, *Graph Autoencoder* (GAE), handles the task of domain adaptation across HSNs while a second component, the *structure-dependent classifier*, ensures the classification task. This design allows the application of fundamentally different classifiers on the top of the same cross-network adapter model.

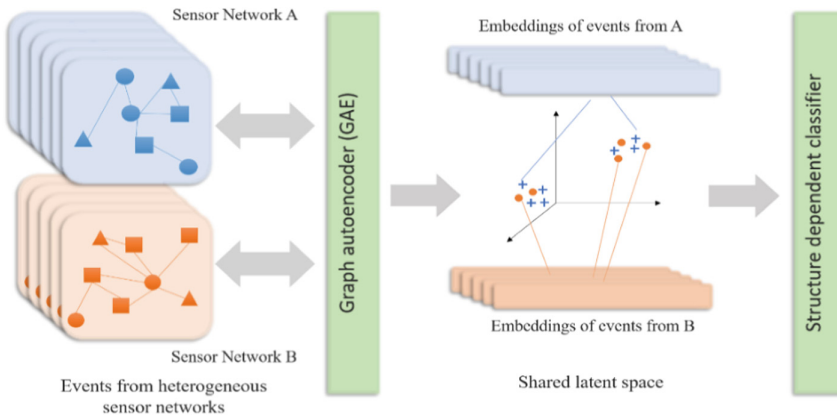


Fig. 1. Cross-network adaptation framework based on a graph autoencoder

We apply the proposed framework to the use case of *activity recognition in smart homes*. Smart homes typically have different layouts and sensors. In practical use cases such as abnormality detection in the routines of solitary elderly people, the well-being of the inhabitant would depend on the deployment of a well-trained model for accurate activity recognition. Focused on this critical use case, we train our encoder with data gathered in CASAS datasets [6] that were collected in smart homes with heterogeneous sensor layouts. The contributions of this paper are as follows:

- We propose a simple graph representation for HSN deployed in smart homes to enable the application of deep learning methods to the HSNs.
- We introduce a novel graph autoencoder architecture, which uses graph convolutional networks [7] together with differentiable pooling [8] to project data coming from HSNs into a latent space with fixed dimensions.

- We leverage associative domain adaptation [9] on our graph autoencoders to minimize the discrepancy of latent representations issued from different sensor networks and thus make the latent space sharable between sensor networks.

The rest of this paper is organized as follows. In Sect. 2, we introduce related methods which have been used for cross-domain adaptation or learning transfer across sensor networks. Section 3 then provides necessary prerequisites and describes the proposed framework. In Sect. 4, we introduce the experimental setup, subsequent parameters settings for our model and collected results. We discuss in Sect. 5 the results and parameters influence. Finally, we conclude and open future directions in Sect. 6.

2 Related Work

Several architectures have been proposed to perform heterogeneous domain adaptation (HDA), which aims at applying to a target domain the knowledge acquired from a source domain. Methods can be divided in two categories: *feature remapping* methods, focusing on finding a mapping between the features in source and target domains, and *latent space transformation* methods, constructing projections of data from source and target domains in a shared latent space.

The key principle of *feature remapping* is to find an optimal mapping between features of heterogeneous domains where the features in target domain can be associated to a single feature or to a combination of features in the source domain. Given the sensor readings of source and target domains and the labels for source domain, Hu and Yang [10] assumed that feature representations in source and target domains are similar. Based on this assumption, they constructed a *translator* that automatically finds a mapping between source and target features. More recently, Feuz and Cook [11] investigated heuristic methods to find an optimal many-to-one mapping through the use of greedy search and genetic algorithms. For these heuristic methods, reported results show greater accuracy and average recall than a manual feature mapping method; nevertheless, the gigantic search space makes the algorithms computationally expensive.

Among those approaches not limited to sensor networks, Zhou et al. [12] formalized an optimization algorithm to learn feature remapping with pivots. While other features are domain-dependents, pivots are described as features that can be commonly shared across domains; therefore, pivots can be used as the guides to transfer domain-specific features. Zhou et al. [13] introduced an algorithm to learn a sparse feature transformation for heterogeneous domain adaptation which allows them to transfer knowledge among support vector machines. Sukhija et al. [14] extended this work using random forests to estimate label distributions that are used as pivots across domains. They compared the models with a baseline of classification task on CASAS datasets [6] and showed a significant decrease of the mean error. However, sparse feature transformation is a supervised method, hence it requires labels from target domain which is often not suitable for real applications.

While *feature remapping* approaches heterogeneous domain adaptation with a direct mapping between domains, *latent space transformation* focuses more on

constructing a projection space shared across domains. The common principle is to learn to (1) project domain-specific data to a specific latent space, (2) use the labels from the source domain to perform a task on the latent space, (3) project data from target domains in the latent space, and finally (4) perform the desired task, e.g. classification, on latent projections. The main challenge resides in the construction of the latent space projection. Shi et al. [15] introduced heterogeneous spatial mapping, an unsupervised method, to align source and target domains using spectral properties, sample selection and Bayesian modelling of output spaces. Another example is Wang et al. [16], which provides an algorithm to construct the projection that preserves the local sample neighborhood in source manifolds while letting similar samples from different manifolds be the neighborhoods in the projection space.

Recent advances in deep learning bring new robust frameworks to approach *latent space transformation*. Zhuang et al. [17] used deep autoencoders to create latent embeddings of domain-specific features. The Kullback-Leibler divergence between embedding distributions is minimized, which allows embedding spaces to converge to a shared space. Wang et al. [18] also leveraged autoencoders to construct a shared feature space, using Maximum Mean Discrepancy and a manifold alignment term to preserve the local geometric structure of data while reducing differences in latent features distributions. Recently, Haeusser et al. [9] introduced *associative domain adaptation*, a method taking advantage of labels in the source domain to create clustered representations of data from source and target domains. This method requires to have labels from the source domain but has the benefit of preserving local structure and thus can create latent spaces with consistent clusters.

In our work, we propose a *latent space transformation* framework based on a new method that performs representation learning on the sensor events modelled as graphs. Representation learning for graphs has been investigated in several works. In particular, Kipf et al. [19] introduced a two kinds of Graph Autoencoders (GAEs) using an inner product between latent variables to reconstruct the adjacency matrix of graphs. Wang et al. [20] leveraged linear graph convolutional networks and GAE to propose a new autoencoder called Marginalized Graph Autoencoder (MGAE). This model corrupts input nodes representation by randomly turning some components to zero. For models using graph convolutional networks to acquire a representation of node features and adjacency, the major issue resides in the variable size of the latent representation, which depends on the number of nodes in the graph. Here, we present a novel GAE architecture that leverages *associative domain adaptation* and *differentiable pooling* to acquire a structure independent, domain invariant representation of graphs.

3 Structure-Independent Graph Autoencoder for HSNs

The structure-independent model introduced here enables usage of ML across HSNs. It relies on graph convolutional networks, differentiable pooling and associative domain adaptation loss to learn to project graph representations into a shared latent space. We first introduce the notation and prerequisite concepts in Sect. 3.1, before presenting our framework in Sect. 3.2.

3.1 Prerequisites

Notations. Let $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{W}\}$ be a graph with a set of N nodes $v_i \in \mathcal{V}$, a set of edges $e = (v_i, v_j) \in \mathcal{E}$, and edge weights $\mathcal{W}(e) \in \mathbb{R}$, $\forall e \in \mathcal{E}$. Given a node ordering for \mathcal{G} , we use the notation $\mathcal{G} = \{X, A\}$, where $A \in \mathbb{R}^{N \times N}$ represents the *adjacency matrix* and each node v_i is represented by a vector of F attributes and collected in a *matrix of node embeddings*, $X \in \mathbb{R}^{N \times F}$. In the following, the notation \tilde{A} is used to refer the *normalized adjacency matrix* $\tilde{A} = \bar{D}^{-1/2} \bar{A} \bar{D}^{-1/2}$, where $\bar{A} = A + I_N$, I_N is the $N \times N$ identity matrix and D is the degree matrix, $\bar{D}_{ij} = \sum_j \bar{A}_{ij}$ if $i = j, 0$ otherwise.

Graph Convolutional Networks. Graph Convolutional Networks (GCNs) [7] learn a *convolution kernel* that uses features of neighboring nodes to perform a potentially non-linear mapping of node embeddings from their initial representation to a different space. A GCN can be compounded of several layers stacked together. Let d_l be the number of features at layer l . In a L -layer(s) GCN, a single layer transforms nodes representation from $\mathbb{R}^{N \times d_l}$ to $\mathbb{R}^{N \times d_{l+1}}$ by applying:

$$X^{(l+1)} = \sigma(\tilde{A}X^{(l)}W^{(l)}) \quad (1)$$

where $W^{(l)} \in \mathbb{R}^{d_l \times d_{l+1}}$ is a matrix of trainable parameters, $\sigma(\cdot)$ is an activation function, $X^{(l)} \in \mathbb{R}^{N \times d_l}$ represents the *matrix of node embeddings* at layer l .

A common application is to use the last layer nodes representation as an input for a differentiable classifier, e.g. a Multi-Layer Perceptron (MLP). The entire model can be efficiently end-to-end trained to classify graph-structured data. In the following, we use the notation $\text{GCN}(A, X) = \tilde{A}XW$ to denote the application of graph convolution on a graph $\mathcal{G} = \{A, X\}$ with a linear activation function.

Differentiable Pooling. As presented in Eq. 1, a layer of a GCN transforms the nodes representation without modifying the graph structure. Ying et al. [8] introduced a differentiable pooling method to modify the graph structure and to learn hierarchical representations of graphs. Pooling layers, also called DIFFPOOL layers, are interleaved between some layers of a graph neural network to coarse the graph representation. A DIFFPOOL layer transforms the original graph structure by clustering nodes. At the l -th layer, we denote the input matrix of embeddings as $X^{(l)} \in \mathbb{R}^{n_l \times d_l}$ and the input adjacency matrix as $A^{(l)} \in \mathbb{R}^{n_l \times n_l}$. The DIFFPOOL layer constructs a new adjacency matrix, $A^{(l+1)} \in \mathbb{R}^{n_{l+1} \times n_{l+1}}$, and a new matrix of embeddings $X^{(l+1)} \in \mathbb{R}^{n_{l+1} \times d_{l+1}}$, such as $(A^{(l+1)}, X^{(l+1)}) = \text{DIFFPOOL}(A^{(l)}, X^{(l)})$.

To perform the transformation between graph structures, a DIFFPOOL layer learns to construct an *assignment matrix* which defines how the nodes in the input graph are *assigned* to the nodes in the output graph. Given $S^{(l)} \in \mathbb{R}^{n_l \times n_{l+1}}$ an assignment matrix for the l -th layer, the transformation applied to the graph structure is:

$$X^{(l+1)} = S^{(l)T} Z^{(l)}$$

$$A^{(l+1)} = S^{(l)T} A^{(l)} S^{(l)}$$

Two GCNs can be used to construct a new matrix of node embeddings $Z^{(l)}$ and the assignment matrix $S^{(l)}$:

$$Z^{(l)} = \sigma\left(\text{GCN}_{l,Z}\left(A^{(l)}, X^{(l)}\right)\right)$$

$$S^{(l)} = \text{softmax}\left(\text{GCN}_{l,P}\left(A^{(l)}, X^{(l)}\right)\right)$$

where $\sigma(\cdot)$ is an activation function and $\text{softmax}(\cdot)$ is the row-wise function defined as $\text{softmax}_i(x) = \frac{\exp(x_i)}{\sum_j \exp(x_{i,j})}$.

We emphasize that both the matrix of node embeddings and the adjacency matrix are constructed using GCNs with fixed output dimensions. Hence, DIFFPOOL layers learn to project into a *latent space with fixed dimensions* the graph representations with potentially different adjacencies or number of nodes.

Associative Domain Adaptation. Introduced by Haeusser et al. [9], associative domain adaptation aims at acquiring consistent projection domains from statistically different source domains. Given Z_i^s and Z_j^t as the embeddings respectively from source and target domains, associative domain adaptation [9] computes a similarity measure matrix using the dot product $M_{i,j} = \langle Z_i^s, Z_j^t \rangle$. Considering the set of source and target representations as a bipartite graph, this similarity measure matrix is used to estimate the transition probability from node embedding Z_i^s to node embedding Z_j^t :

$$P_{ij}^{st} = \mathbb{P}\left(Z_j^t | Z_i^s\right) = \frac{\exp(M_{ij})}{\sum_{j'} \exp(M_{ij'})}$$

From the transition probability, the authors derive two losses which will be minimized during the training step. The first one, *walker loss*, forces round trips within the same class to have the same probability by expressing $\mathcal{L}_{\text{walker}} = H(T, P^{st})$, where H is the cross-entropy function, $P_{ij}^{st} = (P^{st} P^{st})_{ij}$ is the probability of a first order round trip and $T_{ij} = 1/|Z_i^s|$ if i and j have the same class, 0 otherwise. The second loss is a *visit loss* that forces a uniform probability of visiting target examples, $\mathcal{L}_{\text{visit}} = H(V, P^{\text{visit}})$, where $V_j = |Z^t|$ and $P_j^{\text{visit}} = \sum_i P_{ij}^{st}$.

The overall loss $\mathcal{L}_{\text{assoc}}$ is the weighted sum of walker and visit losses:

$$\mathcal{L}_{\text{assoc}} = \beta_{\text{walker}} \mathcal{L}_{\text{walker}} + \beta_{\text{visit}} \mathcal{L}_{\text{visit}} \quad (2)$$

where β_{visit} and β_{walker} are weight factors. A better domain adaptation is achieved if the second weight β_{visit} is decreased when label distribution strongly varies between source and target domains.

3.2 Framework Presentation

The number of sensors, their types and their layouts vary among sensor networks. Intuitively, we want to exploit DIFFPOOL layers ability to build latent spaces of fixed dimensions together with the capacity of associative domain adaptation to acquire domain invariant representations. The key idea is to construct structure-independent representations of HSNs states usable as inputs for structure-dependent ML algorithms. First, we construct a simple graph representation that captures some structural and semantic knowledge about sensor types and layout in a smart home. Next, we present our structure-independent autoencoder. Finally, we introduce the overall objective function enabling domain-invariant representation learning.

Simple Graph Representation for Sensor Networks in Smart Homes. We focus here on simple sensors attached to appliances in smart homes and producing 1-dimensional measures. The minimal semantic information necessary to work with these sensors is their locations in the smart home, e.g. kitchen or bathroom, and the quantity/state they measure, e.g. room temperature or door state. Adjacency between locations is also required. Given semantic information and an ordering for sensors, we define the following rule to build adjacency matrices for smart homes:

$$A_{ij} = \begin{cases} 1 & \text{if sensors } i \text{ and } j \text{ are in the same location} \\ \frac{1}{2} & \text{if } i \text{ and } j \text{ are in adjacent locations} \\ 0 & \text{otherwise} \end{cases}$$

As this design of adjacency matrices is arbitrary, we discuss its pertinence and influence in Sect. 5.3. We also need to build a simple node representation of sensor states at a particular time. We first define an ordered set of generic sensor types representing the quantities or states measured by sensors. As we are informed of sensor state changes through events, we need to gather sensor events to construct features that are representatives of the overall sensor network states at a particular time. We use fixed-window sampling to gather sensor events in non-overlapping windows with duration T_{window} . For a window $[t, t + T_{window})$, we construct the sensor network representation X_t by counting the number of times each sensor fired, i.e. if the i -th sensor fires n times in the window and is of the j -th type, then $(X_t)_{ij} = n$. This guarantees that the features of sensors from different networks are consistently constructed: each line of the matrix X_t is a one-hot vector representing the sensor's type and number of firings within the window.

Encoding the Graph Representation. Like other autoencoders, our model consists of an encoder part, that learns to project the data in the latent space, and a decoder part, that reconstructs input data from the latent representation. The encoder is compounded of at least one differentiable pooling layer which uses node features to perform the

projection in a fixed-size latent space. The latent representation is used as input for the model learning to perform the application task, e.g. classifying states.

For a set of graph representations with the same feature representation but possibly different number of nodes, we construct the latent representations by applying our encoder model, $Z = \text{ENCODER}(A, X)$. For all positive non-zero number of nodes in the graph representation, the matrix of encoded embeddings satisfies $Z \in \mathbb{R}^{N_H \times D_H}$, where N_H and D_H are hyperparameters of the `ENCODER` model. Hence, we construct different graph representations for different sensor networks, and encode them in a latent space of custom dimensions. We implement a simple encoder model defined as follows:

$$\begin{aligned} H_{enc} &= \text{ReLU}(\text{GCN}_H(A, X)) \\ A_{enc}, Z &= \text{DIFFPOOL}(A, H_{enc}) \end{aligned}$$

where the first layer uses the Rectifier Linear Unit, $\text{ReLU}(x) = \max(0, x)$, as activation function. The `DIFFPOOL` layer uses `GCN`, as presented in Eq. 1 with the *hyperbolic tangent* as activation function (denoted $\sigma(\cdot)$ in Eq. 1).

Training the Graph Autoencoder. From the matrix of encoded embeddings Z , the decoder constructs an approximate node representation \hat{X} . As the encoder outputs a graph with N_H nodes $\{A_{enc}, Z\}$, we use the assignment matrix S from the `DIFFPOOL` layer of the encoder to create a representation $H = SZ$ with the same number of nodes as in the input graph. Then, the decoder reconstructs from H an approximate node representation $\hat{X} = \text{DECODER}(H)$. We use the following two-layered decoder model:

$$\begin{aligned} H_{dec} &= \text{ReLU}(\text{GCN}_{dec,1}(A, H)) \\ \hat{X} &= \text{ReLU}(\text{GCN}_{dec,2}(A, H_{dec})) \end{aligned}$$

The overall loss used for training is compounded of several losses with complementary objectives. To train the model as an autoencoder, we define a reconstruction loss \mathcal{L}_{rec} that measures the distance between the input matrix of node embeddings X and the reconstructed representation \hat{X} . In our experiments we used the Euclidian distance between X and \hat{X} , generally known as L2 loss, although many different losses can be used.

A second loss is associated with the assignment matrix of the pooling layer. Ying et al. [8] introduced this loss to ensure that the pooling layer finds a nearly optimal clustering with well-defined nodes assignments. It is expressed as:

$$\mathcal{L}_{pool} = \|A - SS^T\|_F + \frac{1}{n} \sum_{i=1}^n H(S_i)$$

where H is the entropy function and S_i the i -th row of the assignment matrix S .

The last loss is the associative domain adaptation loss from Eq. 2, \mathcal{L}_{assoc} , ensuring the consistency of the shared latent space across domains. The overall loss for the graph autoencoder training step is $\mathcal{L}_{ae} = \mathcal{L}_{rec} + \alpha_{pool}\mathcal{L}_{pool} + \alpha_{assoc}\mathcal{L}_{assoc}$, with the intuitive

recommendation of choosing $\alpha_{\text{pool}}, \alpha_{\text{assoc}} \in [0; 1]$. The set of model parameters is iteratively optimized to decrease this global loss.

4 Experimental Setup

We use the previously introduced encoder model to project HSNs states in a shared latent space. The ML model to transfer across HSNs is then designed and trained on the latent projections made by the encoder model. The overall model, composed of both the encoder and the model to transfer, is trained and used across HSNs. We focus here on using our encoder model to transfer activity recognition models across HSNs deployed in smart homes. Section 4.1 introduces the datasets used in our experimental setup and the preprocessing applied to data. In Sect. 4.2, we present the models and hyperparameters used in our experiments.

4.1 Datasets and Preprocessing

Datasets. We use *hh101* to *hh105*, a set of Human Horizon (HH) datasets collected by CASAS [6] on several smart homes with single occupants. The heterogeneity of layouts among smart homes, depicted in Fig. 2, makes these datasets good data sources to evaluate domain adaptation methods. Each dataset provides raw sensor data with annotated activities. The sensors deployed in the smart homes are either binary sensors attached to appliances or real-valued sensors such as temperature or brightness sensors. Omitting the battery sensors, we count a total of 6 different sensor types. In the dataset, accurate activity annotations result in many labels, which leads to unbalanced class representations. We gather similar classes together to form 13 labels used for activity recognition. In Table 1, we provide a description of the datasets with the number of sensors, events, activities as well as the sensor types and activity labels used for evaluation.



Fig. 2. Example of heterogeneous sensor layouts in the smart homes used to collect *hh101* (left) and *hh103* (right). Different shapes represent different sensor types.

Table 1. Description of Human Horizon datasets from CASAS [6]

Dataset	hh101	hh102	hh103	hh104	hh105
Nb. of sensors	40	64	37	70	53
Nb. of events	321,645	407,583	164,908	478,003	222,591
Sensor types	Door switch, Light switch, Light, Wide area motion, Temperature, Motion				
Activity clusters	Unclassified, Personal hygiene, Cooking, Eating, Working, Entering/Leaving home, House keeping, Taking medicine, Washing dishes, Toilet, Relaxing, Exercising, Other				

Preprocessing. Using smart home layouts as depicted in Fig. 2, we apply the method presented in Sect. 3.2. to create simple graph representations of smart homes. We use a window length of $T_{window} = 180$ s to create graph representations. The activity label associated with a window is set to the activity performed during the longest part of the time window. The six different sensor types used to build representations of sensor states are presented in Table 1.

4.2 Experimental Setup

We use simple and well-known ML algorithms to perform activity recognition: decision trees (DT), k -nearest neighbors (KNN) and multilayer perceptrons (MLP). Our DT classifiers use Gini impurity to determine the quality of splits. We always choose the best split to recursively divide a set of training samples in two subsets, with a maximum number of leaves set to 500. Our KNN models classify a point by returning the value of the $k = 3$ nearest neighbors. We construct two-layered MLP with 64 hidden units and *hyperbolic tangent* activation in the first layer, and 13 output units with *softmax* activation in the second layer. We refer to these baseline models as **DT**, **KNN** and **MLP**. We use a set of preprocessed data with the corresponding activity labels to train three activity classifiers without encoder. Data are divided between a training set (90%) and a testing set (10%). The MLP model is trained to minimize the cross-entropy loss on sample labels, using a batch size of 64 and Adam optimizer [21] with a learning rate of 5×10^{-4} . We add an L2 regularizer loss with a weight of 1×10^{-4} to avoid overfitting.

We used Python 3 and the PyTorch framework to create the GAE model with the architecture described in Sect. 3. The first GCN layer transforms nodes representations from the $F = 6$ shared features, i.e. the number of sensor types, to 32 latent features. Next, the DIFFPOOL layer performs a projection in a latent space of fixed dimensions $N_H \times F_H$, with $N_H = 64$ and $F_H = 16$. We empirically selected appropriate loss weights, resulting in a pooling loss weight $\alpha_{pool} = 1 \times 10^{-4}$ and an association loss weight $\alpha_{assoc} = 0.1$, with $\beta_{walker} = 1$ and $\beta_{visit} = 0.3$. The model is trained with a batch size of 64 using Adam optimizer and a learning rate of 5×10^{-4} . We trained the model on 100 epochs, using early stopping with a patience of 10, i.e. the training stops if the validation loss does not decrease for 10 consecutive epochs. We used two different

datasets to train the autoencoder. Training samples are then encoded with our model and used to train new classifiers with the same hyperparameters as the baselines. We refer to the combination of the GAE with DT (resp. KNN and MLP) models as **GAE+DT** (resp. **GAE+KNN**, **GAE+MLP**). We take advantage of MLP differentiability by end-to-end fine-tuning the GAE+MLP model with a learning rate of 5×10^{-5} .

In the following, we use two metrics to evaluate our models. The first one is the well-known *F1-score*, calculated for a classification task as the harmonic average of precision and recall. We report the average and 95% confidence interval of the F1-score collected on 10 experiments. Due to strong variations between baseline scores, an absolute metric like F1-score is heterogeneous across models. We introduce a second metric, called *relative score*, to compare on a common scale model transferability across datasets. Given a model composed of a classifier trained with a graph autoencoder, the relative score for a target dataset is the ratio between the average F1-score obtained by the model and the average F1-score achieved by the classifier baseline. This score quantifies how the model performs *relatively* to its baseline.

5 Results and Discussion

Our framework aims at applying the same ML model to classify activities using sensor events from several HSNs deployed in smart homes. The activity classifier model can be trained using available labels and data from one or several datasets. We focus on evaluating the performances of our model in a simple case where the activity classifier model is trained on a single *source* dataset, representing data from a *source* smart home layout. The model is used to classify activities on a *target* dataset, representing a *target* smart home layout. Results for transfer of activity classifiers models for different source \rightarrow target datasets pairs are reported and discussed in Sect. 5.1. Due to the use of associative domain adaptation loss, the encoder model requires *data* from the target dataset in addition to *data and labels* from the source dataset. In Sect. 5.2, we quantify the relative score for activity recognition depending on the quantity of available data for target dataset to evaluate how fast activity classifier will adapt to unseen HSNs. In Sect. 5.3, we discuss the pertinence of our arbitrary adjacency matrix design and evaluate the resilience of our framework to suboptimal adjacency matrix design by comparing transfer results for different kinds of adjacency matrix.

5.1 Results for Transfer of ML Models Across HSNs

We evaluate the ability of the encoder to transfer the knowledge acquired on a source HSN to a target HSN. GAE+DT, GAE+KNN and GAE+MLP models are trained using *data and labels* from a source dataset but *only data* from a target dataset. This experiment represents the practical case where labeled data are available only for the source sensor network, but some unlabeled data have been collected on the target sensor network. The results collected for some source \rightarrow target pairs are reported in Table 2.

Table 2. F1-score and relative score of GAE models evaluated on different datasets pairs.

Model		hh101 → hh102	hh102 → hh101	hh103 → hh105	hh104 → hh103	hh105 → hh104
GAE +DT	F1-score	56.6 (0.9)	57.8 (1.7)	62.6 (1.0)	52.3 (1.9)	45.3 (0.6)
	Relative score	83.7	75.7	80.7	75.8	69.4
GAE +KNN	F1-score	55.6 (0.3)	54.5 (0.6)	62.2 (0.6)	50.5 (0.8)	44.1 (0.3)
	Relative score	84.8	74.1	81.6	74.6	70.3
GAE +MLP	F1-score	57.8 (0.7)	57.1 (2.5)	65.1 (0.7)	48.1 (0.9)	45.8 (0.5)
	Relative score	104.7	85.1	92.7	81.3	81.2

Results show that:

- GAE+DT reaches a mean F1-score of 54.9% on the presented source-target pairs. The model performs differently among datasets pairs, with relative score reaching up to 83.7% of the DT baseline score for the pair hh101 → hh102 and down to 69.4% of the DT baseline score for the hh105 → hh104. In average, the GAE+DT model trained on a source dataset achieves 77.0% of the DT baseline score on the target dataset.
- The GAE+KNN model achieves an average relative score of 77.1% of the KNN baseline score, with an average F1-score of 53.4%. Like the GAE+DT model, the higher relative score, 84.8%, is reached for the pair hh101 → hh102. The GAE+KNN model obtains its lower relative scores for hh102 → hh101 and hh105 → hh104, with respectively 74.1% and 70.3% of the KNN baseline score.
- The GAE+MLP model reaches an average relative score of 88.7% of the MLP baseline score, outperforming GAE+DT and GAE+KNN. Once again, the highest relative score is achieved for the pair hh101 → hh102, with a F1-score of 57.8% representing 104.7% of the MLP baseline score. Like GAE+DT and GAE+KNN, the model achieves its lower relative scores for hh104 → hh103 and hh105 → hh104 with respectively 81.3% and 81.2% of the baseline score.

All three models perform differently across source-target pairs. However, we observe that the scores are consistent among models: all of the models achieved their highest relative score on hh101 → hh102 and their second highest relative score on hh103 → hh105. GAE+DT and GAE+KNN reached their two lowest relative score on hh105 → hh104 and hh102 → hh101, while GAE+MLP achieved its lowest relative scores on hh105 → hh104 and hh104 → hh103. These results give the impression that the difficulty of the transfer task varies among source-target datasets pairs, causing different relative scores. There is no explanation on the observed results difference and further investigations would be required to identify the causes underlying this difference.

In our experiments, the GAE+MLP model always outperformed the GAE+DT and GAE+KNN models in terms of relative score. This fact is due to two reasons. First, the MLP baseline is a differentiable model with more hyperparameters than the DT or KNN baselines. Hence, despite our efforts to select appropriate hyperparameters, the MLP baseline is more likely than DT or KNN to achieve suboptimal results. Consequently, the selected hyperparameters can be more appropriate for GAE+MLP than for

the MLP baseline, resulting in a greater relative score for the GAE+MLP model. The second reason behind GAE+MLP high scores is fine-tuning. While the encoder and the classifier are only trained separately for GAE+DT and GAE+KNN, GAE+MLP benefits from an end-to-end fine-tuning of the entire model. Hence, the GAE is optimized to help the MLP achieving better activity recognition results. This fine-tuning step helps GAE+MLP to achieve higher relative scores than GAE+DT or GAE+KNN.

5.2 Evaluation of Deployment Speed

Some data from the target dataset is required to apply associative domain adaptation in our framework. We evaluate how the amount of data collected on the target environment influences the average relative score of our models. Relative scores for different amount of data from target datasets are reported on Fig. 3. As we use a fixed-size window of three minutes, each data point represents three minutes, e.g. 20 data points represent one hour.

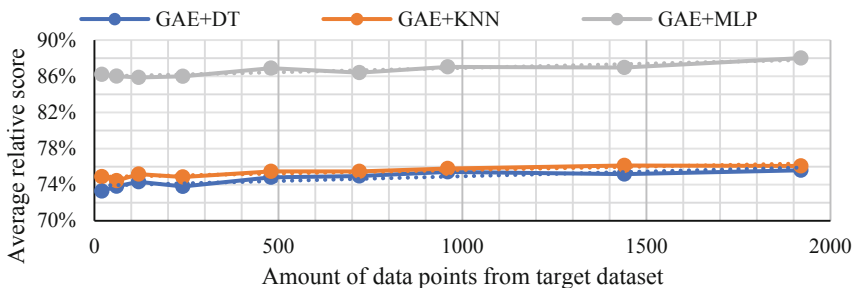


Fig. 3. Evaluation of the adaptation speed: average relative score for different amount of data points from target dataset

We observe that the average relative score will be slightly increased if there is more data from the target HSN. This is the main purpose of using associative domain adaptation. More surprisingly, the model still performs quite decently with a low amount of data from target environment. For instance, the GAE+DT model achieves 73.2% average relative score with only 20 data points from the target dataset, which represents only 3.8% less than the relative score obtained in Sect. 5.1 with every available data points. Hence, the amount of data available from the target HSN has a positive impact on the relative score. Nonetheless, the relative score is decent even with a few data available for the target HSN.

5.3 Influence of the Adjacency Matrix Design

In Sect. 3.2, we arbitrarily design adjacency matrices to express intuitive adjacency between sensors given the layout of an HSN in a smart home. However, this arbitrary design might be suboptimal. Adjacency matrix is a fundamental parameter in graph convolutional layers and strongly influences the way DIFFPOOL layers cluster nodes.

Though it goes beyond our scope to investigate which design of adjacency matrix would best represent sensor layouts in smart homes, we still want to evaluate the influence of the adjacency matrix design.

The previously evaluated model uses the adjacency matrix design from Sect. 3.2, which we call *Default* in the following. The transfer learning results obtained with three additional, different adjacency matrices are compared against the results obtained with our previously evaluated model. The first kind of adjacency matrix (*Identity*) is a zero matrix, which results in the identity matrix after symmetric normalization. The second kind of adjacency matrix (*FC-U*) represents an unweighted fully connected graph. The third kind of adjacency matrix (*FC-W*) is based on the adjacency matrix design introduced in Sect. 3.2 but turned into a fully connected graph by setting the weight between sensor in different non-adjacent locations to 0.1 instead of 0.

For these different adjacency matrix designs, we trained and evaluated GAE+DT models on pairs of source-target datasets. We choose to report results for the GAE+DT model as it achieves decent performances within a limited time and requires only a few hyperparameters tuning. Relative scores obtained for the four different adjacencies designs are presented in Table 3.

We see that the proposed adjacency matrix design outperforms other kinds of adjacency matrices on every source-target dataset pair, except for hh101 \rightarrow hh102. Average relative scores are 77.0% for the *Default* design, 75.0% for the *Identity* design, 74.6% for the *FC-U* design and 75.9% for the *FC-W* design. Hence, if the choice of adjacency in the graph structure helps achieving better transfer results, its influence seems limited. In our case, the worst choice was to represent the sensor network as an unweighted fully connected graph. However, this poor choice of a graph structure only resulted in a loss of 2.4% of the average relative score. We conclude that the influence of the adjacency matrix is limited, and thus our framework is relatively resilient to suboptimal designs of the graph structures representing the HSNs.

Table 3. Relative scores for different kinds of adjacency matrices

Kind of adjacency	hh101 \rightarrow hh102	hh102 \rightarrow hh101	hh103 \rightarrow hh105	hh104 \rightarrow hh103	hh105 \rightarrow hh104
Default	83.7	75.7	80.7	75.8	69.4
Identity	86.5	74.1	76.7	68.0	69.4
FC-U	84.6	72.1	78.9	68.3	69.0
FC-W	83.6	74.3	78.7	75.5	67.4

6 Conclusion and Future Work

When designing an application based on ML and targeting sensor networks, the heterogeneity of sensor layouts and types is a common issue. Indeed, the data collection and model training effort must be repeated for each new sensor network. A practical example is activity recognition in smart homes: a new ML model must be created and trained for each new smart home. In this paper, we propose a new method

leveraging graph representation learning with autoencoders to build latent representations independent from the type or layout of sensors across HSNs.

We introduce a simple graph representation of the state of HSNs in smart home to enable the use of graph autoencoders. Our model relies on differentiable pooling and GCNs to project the representations of events in HSNs to latent spaces of fixed dimension. We then train our model as an autoencoder with an associative domain adaptation term encouraging to share the latent spaces between HSNs. Activity recognition models are trained to classify activities from the shared latent space, which makes the models independent of the structure of target HSNs. We focus here on transferring activity recognition models across HSNs deployed in smart homes. We use CASAS datasets to train baseline activity classifiers models based on decision trees, k-nearest neighbors and multilayer perceptron. After building the graph representation for events in HSNs, we train our graph autoencoder model to create the shared latent representations of events in HSNs. Activity classifiers models are trained on top of the shared latent spaces and compared with their respective baselines.

The obtained results imply that our framework allows to train an activity recognition model based on DT (resp. KNN, MLP) on data from a source smart home with known activity labels and to apply the model to data from a target smart home without activity labels, with an F1-score representing in average 77.0% (resp. 75.0%, 88.7%) of the baseline score on the target smart home. Moreover, the models require only a few data from target HSN to achieve decent performances. In addition, we assume that the proposed adjacency matrix design is suboptimal and we evaluate the influence of the graph structure on the performances of our model. It appears that the structure of the graphs representing the state of HSNs have a quite limited influence on the results. Hence, our model still performs decently with suboptimal graph representations of HSNs.

Enabling the usage of graph neural networks for HSNs opens the range of applicable solutions. The key idea is to instill prior structural knowledge into a ML algorithm, with the intuition that models can exploit this structural knowledge to acquire structure-independent representations of sensor events. However, as we presented in Sect. 3.2, the design of an adjacency matrix, and hence the choice of a structural representation, is a complex but fundamental step. Recent work on graph neural networks by Schlichtkrull et al. [22] focused on instilling semantic knowledge into multi-graphs and proposed frameworks to perform semantic-relation-wise graph convolution. Hence, each layer can acquire a convolution kernel for type of semantic relationship. In these models, adjacency matrices are relation-wise, meaning that we can express different kinds of semantic relationships between sensors. Hence, in the context of cognitive IoT, we can exploit the ubiquity of semantic annotations to create complex structural representations of sensor networks. For instance, sensor adjacency in smart environment could represent that sensors are physically adjacent, but also that they share the same sensor type or are deployed in the same type of room. To summarize, semantic GCNs could take advantage of rich semantic annotations to enhance ML applications.

Acknowledgment. The project reported in this paper is sponsored by Ministry of Science and Technology (MOST) of Taiwan Government under Project Number MOST 107-2218-E-009-020





References

1. Suryadevara, N.K., Mukhopadhyay, S.C., Wang, R., Rayudu, R.K.: Forecasting the behavior of an elderly using wireless sensors data in a smart home. *Eng. Appl. Artif. Intell.* **26**, 2641–2652 (2013). <https://doi.org/10.1016/J.ENGAPPAL.2013.08.004>
2. Orpwood, R., Adlam, T., Evans, N., Chadd, J., Self, D.: Evaluation of an assisted-living smart home for someone with dementia. *J. Assist. Technol.* **2**, 13–21 (2008)
3. Lotfi, A., Langensiepen, C., Mahmoud, S.M., Akhlaghinia, M.J.: Smart homes for the elderly dementia sufferers: Identification and prediction of abnormal behaviour. *J. Ambient Intell. Humaniz. Comput.* **3**, 205–218 (2012). <https://doi.org/10.1007/s12652-010-0043-x>
4. Barth, J., et al.: Biometric and mobile gait analysis for early diagnosis and therapy monitoring in Parkinson’s disease. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 868–871. IEEE (2011)
5. Chiang, Y., Lu, C.-H., Hsu, J.Y.-J.: A feature-based knowledge transfer framework for cross-environment activity recognition toward smart home applications. *IEEE Trans. Hum. Mach. Syst.* **47**, 310–322 (2017). <https://doi.org/10.1109/THMS.2016.2641679>
6. Cook, D.J., Crandall, A.S., Thomas, B.L., Krishnan, N.C.: CASAS: a smart home in a box. *Comput. (Long. Beach. Calif.)* **46**, 62–69 (2013). <https://doi.org/10.1109/mc.2012.328>
7. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. *arXiv Prepr. arXiv1609.02907*. (2016)
8. Ying, R., You, J., Morris, C., Ren, X., Hamilton, W.L., Leskovec, J.: Hierarchical graph representation learning with differentiable pooling. In: *Advances in Neural Information Processing Systems*, pp. 4800–4810 (2018)
9. Haeusser, P., Frerix, T., Mordvintsev, A., Cremers, D.: Associative domain adaptation. In: *Proceedings of IEEE International Conference of Computer Vision*, October 2017, pp. 2784–2792 (2017). <https://doi.org/10.1109/iccv.2017.301>
10. Hu, D.H., Yang, Q.: Transfer learning for activity recognition via sensor mapping. In: *IJCAI International Joint Conference on Artificial Intelligence*, pp. 1962–1967 (2011)
11. Feuz, K.D., Cook, D.J.: Heterogeneous transfer learning for activity recognition using heuristic search techniques (2014)
12. Zhou, G., He, T., Wu, W., Hu, X.T.: Linking heterogeneous input features with pivots for domain adaptation. In: *Twenty-Fourth International Joint Conference on Artificial Intelligence*, pp. 1419–1425 (2015)
13. Zhou, J.T., Tsang, I.W., Pan, S.J., Tan, M.: Heterogeneous Domain Adaptation for Multiple Classes (2014)
14. Sukhija, S., Krishnan, N.C., Singh, G.: Supervised Heterogeneous Domain Adaptation via Random Forests. In: *International Joint Conferences on Artificial Intelligence*. pp. 2039–2045 (2016)
15. Shi, X., Liu, Q., Fan, W., Yu, P.S., Zhu, R.: Transfer learning on heterogenous feature spaces via spectral transformation. In: *Proceedings of IEEE International Conference of Data Mining, ICDM*, pp. 1049–1054 (2010). <https://doi.org/10.1109/icdm.2010.65>
16. Wang, C., Mahadevan, S.: Manifold alignment without correspondence (2009)
17. Zhuang, F., Cheng, X., Luo, P., Pan, S.J., He, Q.: Supervised representation learning: transfer learning with deep autoencoders. In: *IJCAI International Joint Conference on Artificial Intelligence*, pp. 4119–4125 (2015)
18. Wang, X., Ma, Y., Cheng, Y., Zou, L., Rodrigues, J.J.P.C.: Heterogeneous domain adaptation network based on autoencoder. *J. Parallel Distrib. Comput.* **117**, 281–291 (2018). <https://doi.org/10.1016/j.jpdc.2017.06.003>
19. Kipf, T.N., Welling, M.: Variational Graph Auto-Encoders (2016)

20. Wang, C., Pan, S., Long, G., Zhu, X., Jiang, J.: MGAE: marginalized graph autoencoder for graph clustering. <https://doi.org/10.1145/3132847.3132967>
21. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization (2014)
22. Schlichtkrull, M., Kipf, T.N., Amsterdam pbloem, V., Rianne van den Berg, vunl, Titov, I., Welling, M.: Modeling Relational Data with Graph Convolutional Networks Peter Bloem. In: European Semantic Web Conference. pp. 593–607 (2017)



Development of an Acoustically Adaptive Modular System for Near Real-Time Clarity-Enhancement

Alexander Liu Cheng^{1,2} , Patricio Cruz³ ,
Nestor Llorca Vega^{2,4} , and Andrés Mena² 

¹ Faculty of Architecture, Delft University of Technology,
Delft, The Netherlands

a.liucheng@tudelft.nl

² Faculty of Architecture and Engineerings, Universidad Internacional SEK,
Quito, Ecuador

³ Faculty of Electrical and Electronic Engineering, Escuela Politécnica Nacional,
Quito, Ecuador

⁴ School of Architecture, Universidad de Alcalá, Madrid, Spain

Abstract. This paper details the development of an acoustically adaptive modular system capable of enhancing Speech Clarity (C_{50} Clarity Index) in specific locations within a space in near real-time. The mechanical component of the system consists of quadrilateral, truncated pyramidal modules that extend or retract perpendicularly to their base. This enables said modules (1) to change in the steepness of the sides of their frustum, which changes the way incoming sound waves are deflected/reflected/diffused by the surfaces of the pyramid; and (2) to reveal or to hide the absorbent material under each module, which enables a portion of incoming sound waves to be absorbed/dissipated in a controlled manner. The present setup considers a fragmentary implementation of six modules. The behavior of these modules is determined by two steps in the computational component of the system. First, the initial position of the modules is set via a model previously generated by an evolutionary solver, which identifies the optimal extension/retraction extent of each of the six modules to select for individual configurations that collectively ascertain the highest clarity in said specific locations. Second, a simulated receiver at the location in question measures the actual clarity attained and updates the model's database with respect to the configuration's corresponding clarity-value. Since the nature of acoustics is not exact, if the attained measurement is lower than the model's prediction for said location under the best module-configuration, but higher than the second-best configuration for the same location, the modules remain at the initial configuration. However, if the attained values are lower, this step reconfigures the modules to instantiate the second—or third-, fourth-, etc.—best configuration and updates the model's database with respect to the new optimal module-configuration value. These steps repeat each time the user moves to another specific location. The objective of the system is to contribute to the intelligent and intuitive Speech Clarity regulation of an inhabited space. This contributes to its *Interior Environmental Quality*, which promotes well-being and quality of life.

Keywords: Cyber-physical systems · Adaptive acoustics · Internet of Things

1 Introduction and Motivation

This paper details the development of an acoustically adaptive modular system capable of enhancing Speech Clarity (C_{50} Clarity Index) in specific locations within a space in near real-time. It is designed and implemented as a sub-system within an open-ended and on-going development of an intelligent built-environment framework informed by both technical and architectural considerations. With respect to the technical, the present work is situated within the *Ambient Intelligence* (AmI) [1, 2]/*Ambient Assisted Living* [3]—or *Active and Assisted Living* [4, 5]—(AAL) discourse. With respect to the architectural, it is informed by the *Interactive Architecture* [6] and *Architectural Robotics* [7] discourses. The consideration of both aspects is central to the development of mutually complementary interoperability between physical and computational components within the built-environment.

The acoustically adaptive modular system is designed to improve the acoustic ambience via said enhancement of Speech Clarity and a concomitant noise-reduction in predetermined locations via mutually informing Physical/Mechanical and Computational components. Since sound is a potential environmental stressor associated with a variety of negative physiological, psychological, and cognitive responses [8], acoustic ambience is an important indicator of *Interior Environmental Quality* (IEQ) [9], which is strongly correlated with well-being and sustained quality of life [10]. The impact extends to a variety of programmatic functions as well as to specific spaces and audiences, not all partial to the context or character of AmI/AAL. For example, with respect to classrooms and children: Speech Clarity is strongly correlated with reading development among elementary school—i.e., second-grade—pupils [11]; and with respect to offices and adults: it is strongly correlated with intelligibility even in tele-conference systems at the workplace [12], etc.

Although the scope of the detailed implementation consists of maximal C_{50} value-selection at octave band mid-frequency of 500 Hz, the same method and system may be used to select for C_{80} —Music Performance Clarity Index—or other acoustic features such as Reverberation Time, Definition (D_{50}), Early Decay Time, etc., at a variety of frequencies (e.g., 1000 Hz–8000 Hz). Accordingly, the present work is an instance-implementation of a method-type capable of enhancing acoustic ambience with respect to multiple acoustical parameters (in individual maximization or collective optimization). As with other sub-system developments belonging to the same open-ended intelligent built-environment framework, the present system is intended to operate intuitively, intelligently, and automatically in a closed-loop via inattentive or passive user-interaction yet without his/her intervention.

The system is presented in five sections. Section 2 describes the *Concept and Approach*, which explains the reasoning behind the physical as well as computational mechanisms and their interrelation. Section 3 details the *Methodology and Implementation*, which describes the actual implementation of a proof-of-concept fragment consisting of six modules. Section 4 presents the *Results*, which demonstrate the

successful operation of the system as corroborated by a performance overview. Finally, Sect. 5 provides a *Conclusion* and discusses present limitations and future work.

2 Concept and Approach

The present setup considers a virtual space of 4 m in length, 1.5 in width, and 1 in height—N.B.: the width and height correspond to the dimensions of the six-module fragment, and represent minimal dimensions for trials in the present setup. In this volume, the simulated sound-source is placed 3 m from the acoustical six-module fragment along the center-axis of the volume's length. The simulated receiver, and therefore the particular location where maximal C_{50} value is being selected for, is placed at 1 m from the module fragment along the same axis. The virtual and physical acoustic modules measure 0.5×0.5 m (see Fig. 1; see Table 1 for *Absorption Coefficients*). The physical modules instantiate extension/retraction configurations corresponding to C_{50} calculations from this virtual space.

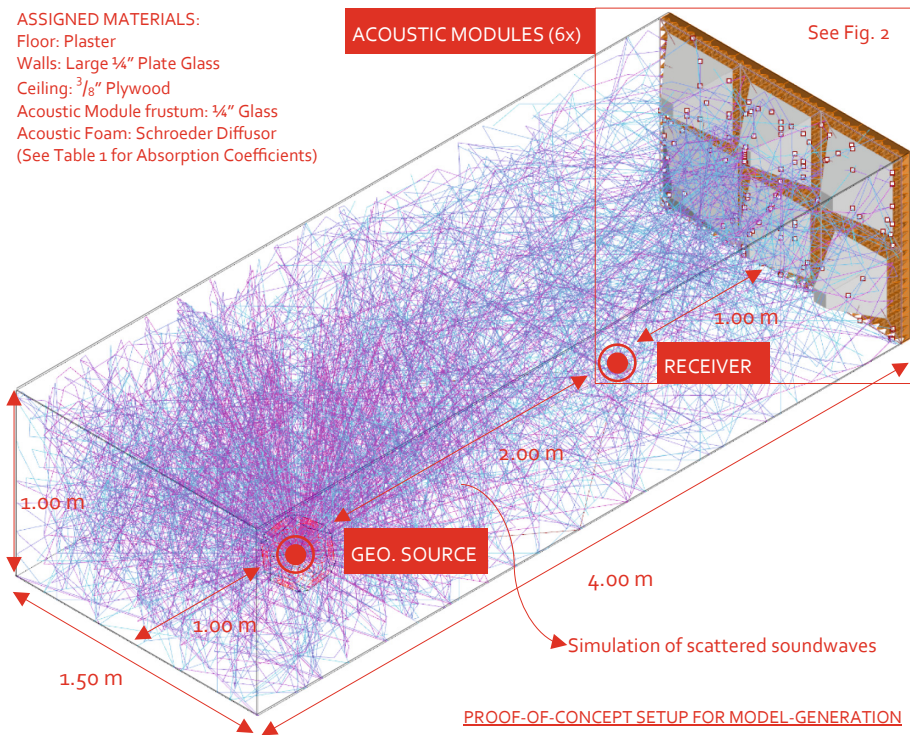


Fig. 1. Sample arrangement of six acoustically adaptive modules. Present configuration attains ~ 5.72 dB at a designated point in octave band mid-frequency of 500 Hz. Sound waves are represented in lines up to ten surface-bounces; the line colors are correlated with the sound-energy (i.e., darker to lighter equals more to less energy, respectively).

Table 1. Absorption coefficients (% energy absorbed)—from Pachyderm [16] material library

Hz	Plaster – rough on lath (<i>Floor</i>)	Large ¼” Plate Glass (<i>Pyramid Surfaces, Walls</i>)	¾” Plywood Wall (<i>Ceiling</i>)	Schroeder Diffusor (<i>Acous. Absorber</i>)
62.5	2	25	32	18
125	2	18	28	22
250	3	6	22	24
500	4	4	17	32
1000	5	3	9	23
2000	4	2	10	19
4000	3	2	11	19
8000	2	2	13	19

The adaptive modular system is conceived as a *Cyber-Physical System* (CPS) [13, 14]. Its physical/mechanical component (see Sect. 3.1) consists of quadrilateral, truncated pyramidal modules that extend or retract perpendicularly to their base (see Figs. 2 and 4). This enables said modules (1) to change in the steepness of the sides of their frustum, which changes the way incoming sound waves are deflected/reflected/diffused by the surfaces of the pyramid; and (2) to reveal or to hide the absorbent material under each module, which enables a portion of incoming sound waves to be absorbed/dissipated in a controlled manner. The extension/retraction of each module—ranging from 70 mm to 270 mm in height—is controlled by the computational component of the system (see Sect. 3.2), which consists of two steps.

In the first, the initial extension/retraction of each module is determined by a generated model based on an evolutionary solver—viz., Galapagos [15]—selecting for maximal C_{50} values ascertained via an acoustical simulation software—viz., Pachyderm [16]—both running on Grasshopper [17]. In the second step, a simulated receiver at the location in question measures the clarity attained and updates the model’s database with respect to the configuration’s corresponding clarity-value. Since the nature of acoustics is not always exact, if the attained measurement is lower than the model’s prediction for said location under the best module-configuration, but higher than the second-best configuration for the same location, the modules remain at the initial configuration. However, if the attained values are lower, this step reconfigures the modules to instantiate the second—or third-, fourth-, etc.—best configuration and updates the model’s database with respect to the new optimal module-configuration value (see Fig. 3). These steps repeat each time the user moves to another location.

The computational model is an open-ended and closed-loop mechanism that is generated before any actual operation of the physical/mechanical component. It is open-ended in that, via its evolutionary solver (see Sect. 3.2), it continues to compute selected module extension/retraction configurations without a specific value as its selected target. It therefore continues to build its database of C_{50} values with respect to module extension/retraction configurations indefinitely. A caveat: the solver may be configured to end either after a particular period of time or when the difference between maximal values found becomes smaller than some relevant threshold—e.g., when the

Octave band mid-frequency: 500Hz (... now selecting for highest C50)
 A-Weighted Sound Pressure Level: 123.384464
 C50, dB: 5.715291
 C80, dB: 9.746634
 Reverb. time: 0.597964
 D50: 78.851451
 Early Decay Time: 0.485165

F	E	D
206.0	224.0	213.0
139.0	230.0	256.0
C	B	A
(mm EXTENSION)		

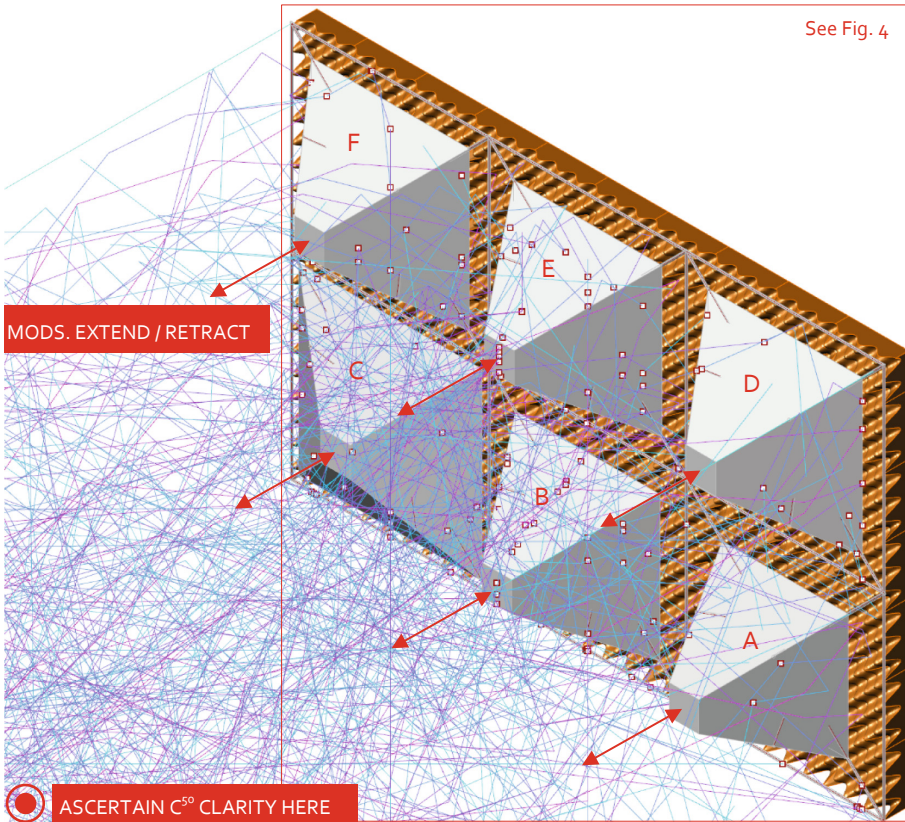


Fig. 2. Arrangement of six acoustically adaptive modules. Present configuration attains ~5.72 dB at a designated point in octave band mid-frequency of 500 Hz.

differences between values are found after ten decimal points, etc. Moreover, in cases where the *Fitness Landscape* or *Volume* (i.e., every fitness value resulting from the inter-combination of different variables or *genes*—see Rutten’s discussion [18]) does have an actual optimum (either maximal or minimal value), the solver would end after having found it. However, this is not the case in the present system, as due to its complexity, there may be several equally satisfactory values whose difference is negligible (again, e.g., values with differences after ten decimal points).

The model is also closed-loop in that it updates its database from received feedback. That is, while the solver computes values from module extension/retraction

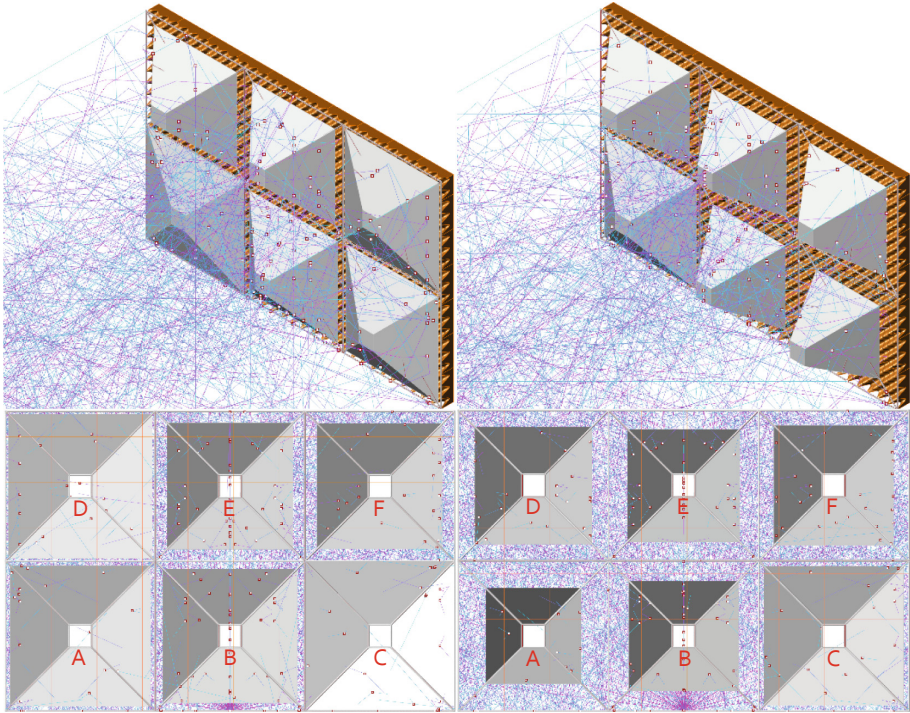


Fig. 3. Top-row: frontal isometric view. Bottom-row: posterior elevation view, viewing towards sound-source—N.B.: acoustic foam transparent). Left-column: iteration #20, with a C_{50} value of 3.8514 dB—acoustic module configurations (values in mm): A. 132; B. 159; C. 85; D. 112; E.187; F. 189 extension. Right-column: iteration #784, with a C_{50} value of 5.715291 dB—acoustic module configurations: A. 256; B. 230; C. 139; D. 213; E.224; F. 206 extension. Note that even small module configuration variations yield vastly different clarity values.

configurations and ranks them from highest (most optimal) to lower (least optimal) indefinitely, the feedback mechanism updates the model's ranked database to reflect actual states of affairs—i.e., the actual supersedes the predicted/generated). In this manner, the model is constantly and continuously ascertaining the latest values and ranking them for use, which enables the physical/mechanical component to operate as soon as the model's database has at least one predicted/generated value. This is why a model extension/retraction configuration is said to be the most optimal with respect to maximizing C_{50} only up until the most recent iteration. Of course, an arguable minimum number of iterations (and therefore stored C_{50} values) is required in order to yield non-trivial module extension/retraction configuration suggestions—that is, to say that a particular module extension/retraction configuration is the optimal because there is only one generated/measured C_{50} value is to say nothing at all. While this setup risks yielding trivial results when the model's database is small, it also enables the system to potentially improve over time (see Sect. 5 for a caveat) with increasingly higher C_{50} values found in a more comprehensive database (for example, see Fig. 6 for a *Histogram* corresponding to presently computed C_{50} values).

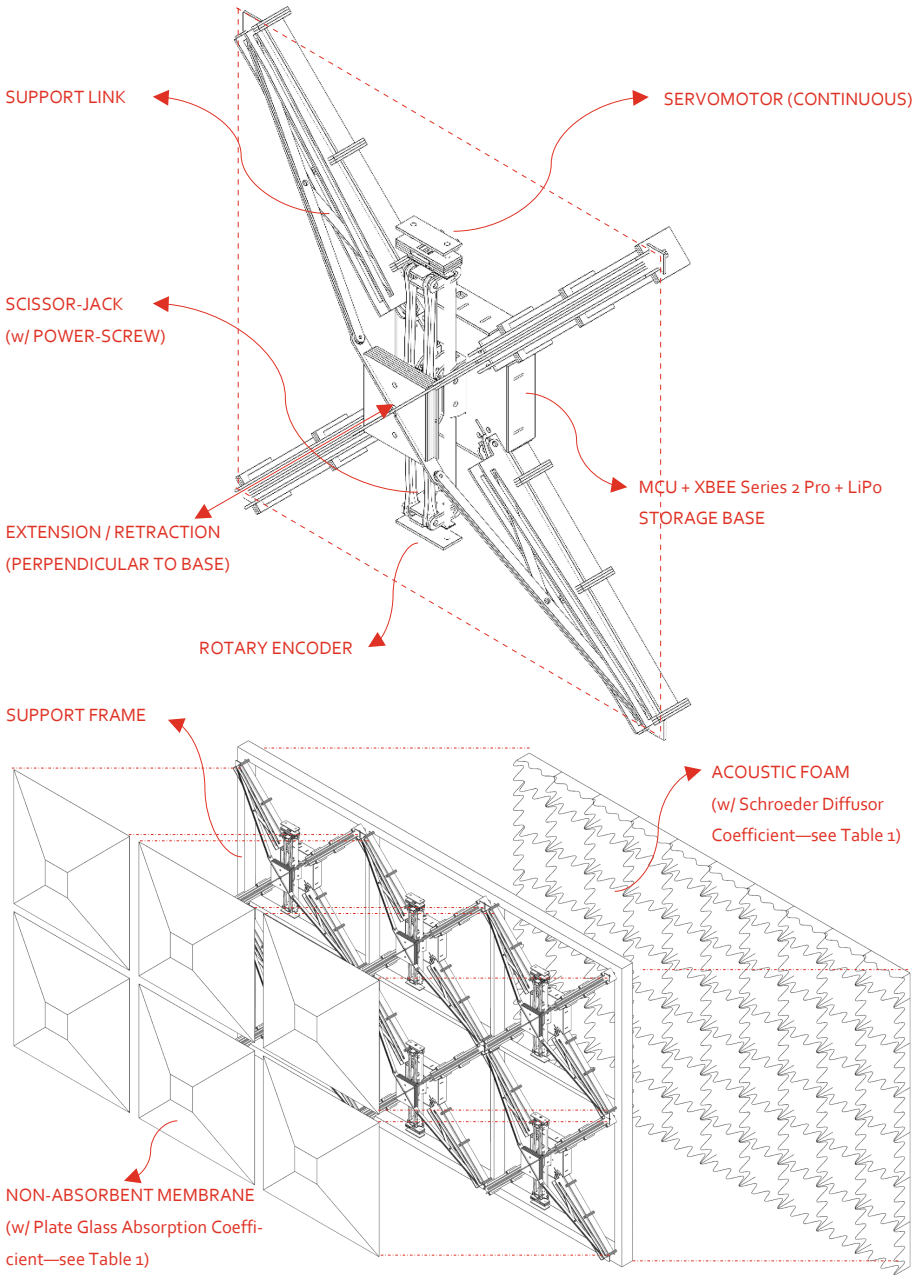


Fig. 4. Top: module breakdown. Bottom: implemented six-module fragment breakdown.

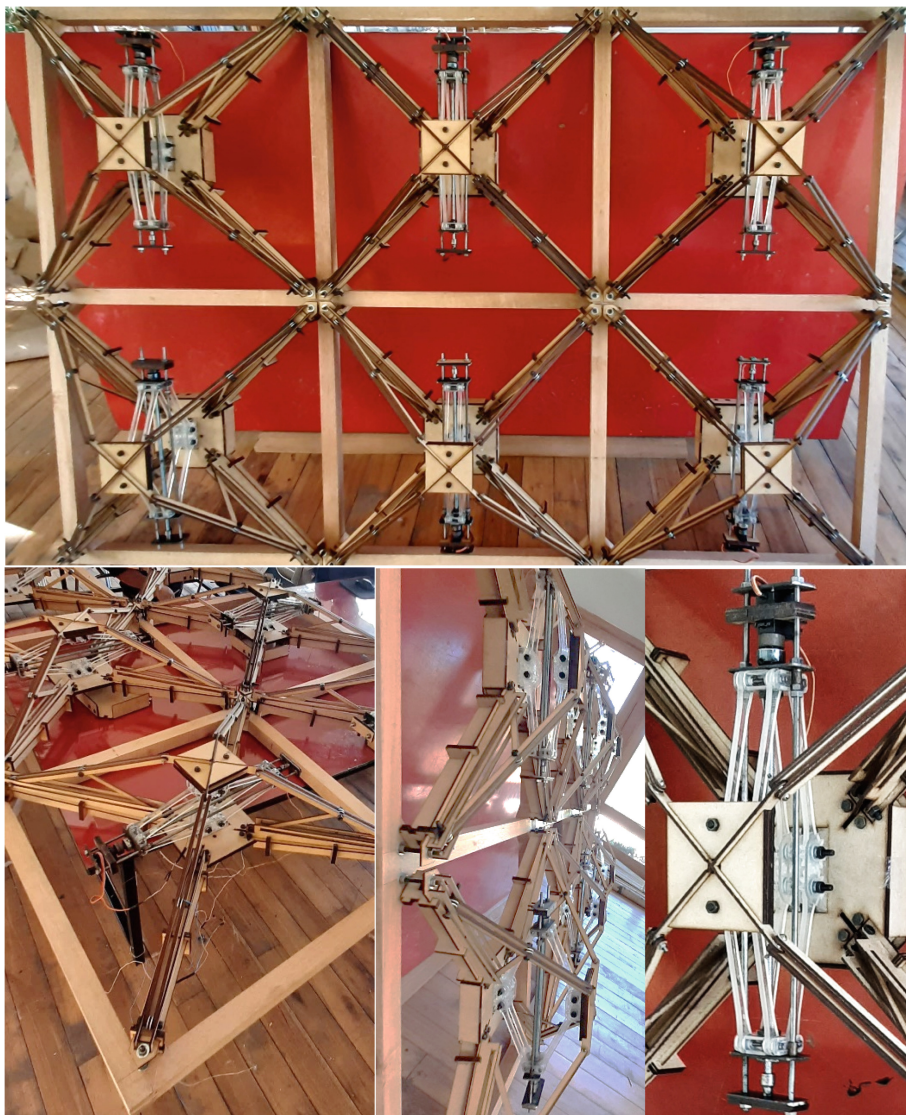


Fig. 5. Physical TRL-4/5 implementation of the six-module fragment used in laboratory tests.

3 Methodology and Implementation

The behavior of the physical acoustic modules with respect to their extension/retraction extents is determined by data gathered from the virtual space. Although the ascertained module configurations are expressed in the physical world, the real space is not correlated with the virtual one. That is to say, in the scope of the present implementation, the principal purpose is to demonstrate that a model configuration found via a virtual

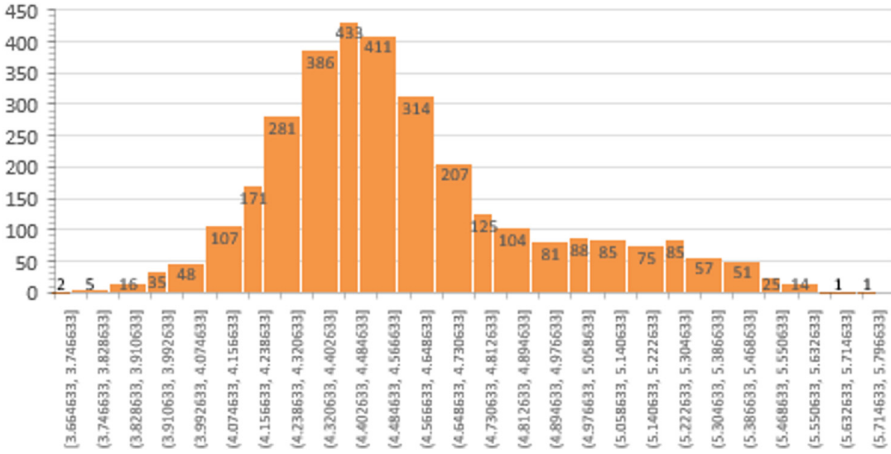


Fig. 6. Histogram of C_{50} values across 3,625 iterations generated in the present implementation.

space may be instantiated in the real-world; and that feedback corresponding to the real-world may be used to update the model’s database. The particulars of both physical and computational components are detailed in the following sub-sections.

3.1 Physical/Mechanical Component: Acoustically Adaptive Modules

In terms of the physical build, each instance of the acoustically adaptive module is built with medium-density fiberboards, a steel tie-rod or power-screw, an acrylic scissor-jack as well as a variety of other metallic accessories such as ball-bearings, bolts, nuts, etc. These modules were designed as proof-of-concept instances appropriate to a *Technology Readiness Level* (TRL) [19] of 4–5 (see Fig. 5).

In terms of *Information and Communication Technologies* (ICTs), each module is equipped with a continuous servomotor (with 15 kg of torque) correlated with a rotary encoder to keep track of the rotation of the power-screw. The motor and encoder are connected to a *Microcontroller Unit* (MCU) attached at the base of the module. An XBee Series 2 Pro antenna as well as a LiPo battery are attached to the MCU—N.B.: While the modules depend on the main power supply to function, batteries are integrated as a contingency measure in case of power-outage. Each module serves as a node in a self-healing and meshed *Wireless Sensor and Actuator Network* (WSAN) whose System Architecture is inherited from previous developments by some of the authors [20]. Each node sends and receives data to and from any one of several coordinating devices built with *Raspberry Pi Zero Ws* (RPiZWs). The computational model stores calculated C_{50} data in a shared database that the RPiZWs access via ZigBee in order to read module extension/retraction extents corresponding to each C_{50} value, which are then relayed—also in ZigBee—to each adaptive acoustic module for configuration instantiation. The present setup considers three coordinating RPiZWs as representative of a larger multitude in an intelligent built-environment’s WSAN. This redundancy ensures system resilience, as operation would not be interrupted if one or

two coordinating nodes were to fail. Moreover, even if the selected communication protocol (i.e., ZigBee, for reduced energy-consumption) between the nodes were to fail, the sending and receiving of data and instructions would default to *User Datagram Protocol* (UDP) over WiFi. Redundancy of computational resources as well as communication protocols is essential to instantiate unobtrusive, intuitive, and independent data-driven intelligence. Finally, the LiPo battery is included as a secondary source of power, as the present implementation presupposes uninterrupted power availability from the main power supply.

3.2 Computational Component: Evolutionary Solver

As mentioned in Sect. 2, the computational component has two steps, where the first instructs physical modules to instantiate a particular extension/retraction configuration; and the second receives feedback and updates the model's database. Both steps are described in greater detail below. But it is worth emphasizing first that prior to either step, the spatial and material attributes and conditions of the virtual space must be assigned and/or determined. That is, the virtual space's (see Fig. 1) walls, ceiling, floor, etc., must be assigned material absorption attributes (see Table 1). Likewise, the sound-source and -receiver must also be explicitly specified and identified. From this, the acoustic simulation software—viz., Pachyderm [16]—may be used to compute a corresponding *Energy-Time Curve* (ETC) from which a variety of acoustical parameters may be ascertained. In the present case, C_{50} is the parameter of interest, which is an objective measure of Speech Clarity (vis-à-vis C_{80} , a measure of Music Performance Clarity) measured in *decibels* (dB). C_{50} , as a Clarity Index, reflects the fact that late sound-reflections degrade the intelligibility of speech due to the merging speech sounds. Similarly, very early sound-reflections, while not detrimental *per se*, will invariably contribute to intelligibility. The time-limit before sound-reflections become detrimental is agreed to be approximately 50 ms. C_{50} is calculated accordingly:

$$C_{50} = 10 \log_{10} \frac{\int_0^{50} p^2(t) dt}{\int_{50}^{\infty} p^2(t) dt}, \quad (1)$$

where $p(t)$ is the impulse-response sound-pressure at time t measured from direct-sound arrival [12]. The present implementation ascertains the highest C_{50} value possible in one octave band frequency: 500 Hz—although the same methods work for other bands.

In the first step of the computational component, the evolutionary solver—i.e., Galapagos [15]—instantiates an initial set of random module extension/retraction configuration in the virtual space and their corresponding C_{50} values are derived via the calculated ETC in Pachyderm [16]. The extension/retraction extents corresponding to the highest C_{50} gathered in this random set are relayed to the MCU of each corresponding module—respectively—and instantiated in the real world. While the physical instantiation takes place, the evolutionary solver continues to gather more C_{50} values from different configurations and ranks them from highest (most optimal) to lowest (least optimal). Only when a recent value is higher than the previously instantiated are new module extension/retraction extents relayed for physical instantiation. This means that given a specific location within a space, the module configurations will continue to

update in near-real time, always instantiating the most optimal value found thus far for that location. Since it is not the case that each iteration in the virtual space yields a progressively higher value, the module configuration will not be changing at each iteration. Instead, the most optimal module configuration thus far for that location will be instantiated automatically and in near real-time whenever the user steps in that location without the need of calculation—i.e., calculation takes place continuously in the background, at every iteration, but there is always a most optimal stored value corresponding to said location for fast physical instantiation.

As previously stated, the first set of module configurations is random. However, from that point onwards the extension/retraction extent of each module proceeds with evolutionary principles from information gathered from said random configurations. That is, each configuration contains an extension/retraction extent value for each module A, B, C, D, E, and F (see Fig. 2). This set or array of values per each configuration is called a *genome*. The *fitness*—in the present case: the C_{50} value—is determined for each *genome*, and the *Fitness Volume* is populated with these values. The fittest *genomes* are bred with one another to create a subsequent iteration or generation. As Rutten [18] points out, this breeding among the fittest *genomes* is necessary as it is unlikely that the initial randomized set instantiates the most optimal solution. Prior to the breeding process, those *genomes* deemed to be unfit—via a configuration option—are discarded and only the set of best performing pass their genes to the next generation. The offspring of the first randomized generation will have *genomes* whose *fitness* is distributed somewhere between that of their parents. A caveat cum limitation: in simpler systems—say, those with only two variables—it may be reasonable to assume that the offspring of fit parents will enjoy a relatively similar degree of fitness. However, this is not necessarily the case in the present six-variable system, where any one variable may have a negative impact over the positive selection of the whole. This *Fitness Volume* may be deemed chaotic due to its unpredictability. However, despite the individual chaos, rhyme and reason may still be derived when considering the sum average of all attained C_{50} values. That is, the *genomes* from each iteration/generation may yield *fitness* values not conforming to a discernible pattern, but the system may be said to be evolving positively if the sum average of said *fitness* values is said to be on the overall increase. In other words, the system is still evolving positively if, on average, subsequent generations are fitter than previous ones, which is the case in the present implementation (see Fig. 7)—see Sect. 5 for an expansion on this caveat with respect to limitations. Despite such potential limitations, evolutionary solvers are still considerably useful in problems that select for maximization or minimization of values. They are, for example, markedly more effective than trying all possible variable combinations in this six-variable system, which would require 64,000,000,000,000 calculations (i.e., the extension/retraction range of 200 mm of each module to the power of 6 modules).

In the second step of the computational component, after having physically instantiated a highest C_{50} value, an independent simulated receiver (also implemented with Pachyderm [16], yields a “measured” C_{50} value to compare with the model-computed value. Since the nature of acoustics is not always exact, simulated receivers may yield different C_{50} values under the same conditions. (N.B.: Even the same simulated receiver may yield a different C_{50} value if measured multiple times (see, for

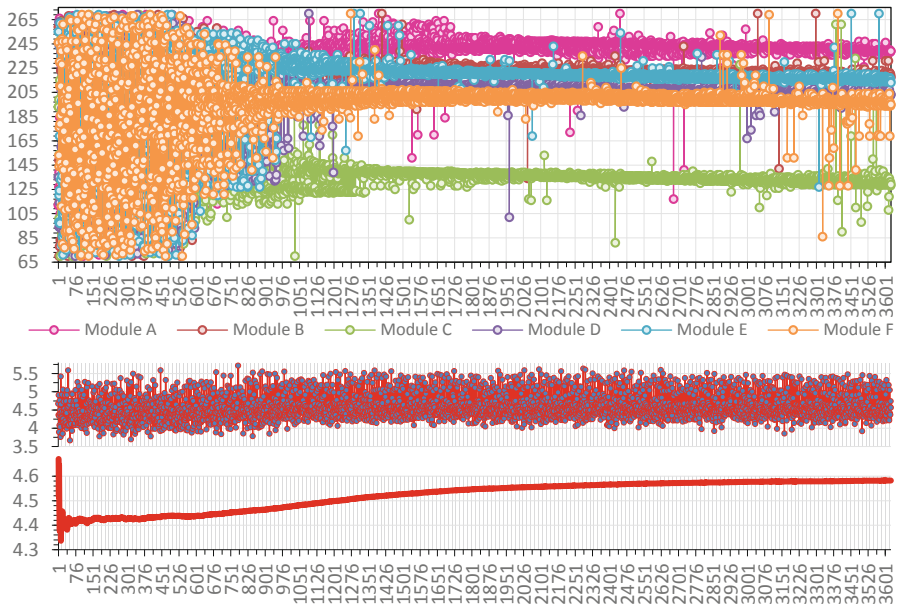


Fig. 7. Top: extension/retraction of each module with respect to each iteration. Middle: Corresponding C_{50} values for each *genome* with respect to each iteration. Bottom: Sum averages across all iterations.

example, iterations # 1 and 2 in Table 2, where with the same module extension/retraction configurations yielded 4.67 dB vs. 4.36 dB, respectively). But said values will also not be substantially different from each other. That is to say, the nature of acoustics is compatible with margins and tolerances.) If the independent receiver's C_{50} value is higher than that of the model's for that particular module configuration, the model's database is updated to reflect this independently measured value and the physical module configuration remains unchanged. If, however, the independent receiver's value is lower than the model's (i.e., lower than expected), then the system updates the model's database and proceeds to look for another module configuration that is expected to be higher than the independently measured value. That is to say, suppose that a given module configuration that corresponds to the highest ranked C_{50} value is expected to ascertain 5.00 dB at a given location but instead is independently measured to yield only 4.5 dB. At this point, the 5.00 dB in the model's database is replaced by 4.5 dB, and the system finds the second highest ranked C_{50} value and compares it to the 4.5 dB. If the second highest ranked C_{50} value is higher—say, 4.75 dB—then the system first proceeds to instantiate the module extension/retraction configuration that corresponds it and then to classify it as the new highest ranked C_{50} value. Hence what used to be the highest ranked C_{50} value is now the second (4.5 dB) and what used to be the second is now the first (4.75 dB). If, however, the second highest ranked C_{50} value is lower than the measured 4.5 dB value—say, 4.25 dB—then 4.5 dB and its corresponding module configuration remain highest ranked and

4.24 dB remains second highest. By virtue of the ranking mechanism, it is impossible for the third highest ranked to be higher than 4.25 dB, 4.5 dB, or 4.75 dB in this example. In other words, the second step of the computational component ensures that the independently measured C_{50} values are integrated back into the model's database always in a ranked manner, where the highest ranked is always the best module extension/retraction configuration to yield the highest C_{50} value for a specific location. This two-step process repeats at every configuration instantiation that yields a C_{50} value that is higher than the previous highest ranked value.

Table 2. Sample module configurations and resulting C_{50} values and sum averages.

Iter.	Mod. A (mm)	Mod. B (mm)	Mod. C (mm)	Mod. D (mm)	Mod. E (mm)	Mod. F (mm)	Output C_{50} (dB)	Σ Average (dB)
1	112	259	193	135	255	246	4.670331	4.670331
2	112	259	193	135	255	246	4.362115	4.516223
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
20	132	159	85	112	187	189	3.8514	4.396181
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
784	256	230	139	213	224	206	5.715291	4.455165
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3625	239	218	129	203	213	195	4.563596	4.581994

4 Result

The computation component of the system executed 3,625 iterations in the present implementation. During this cycle, a maximum C_{50} value of ~ 5.72 dB (at iteration # 784, see Table 2) and minimum of ~ 3.66 dB are ascertained, with the sum average of all iterations being ~ 4.58 dB. The difference between maximum and minimum is ~ 2.05 dB.

The majority of module extension/retraction configurations yielded C_{50} values between ~ 4.40 dB and ~ 4.48 dB (see Fig. 6). This illustrates a salient advantage of implementing adaptive intelligence in the present system, as random manual operation would likely yield results in that range, which is lower than ~ 5.72 dB. At 3,625 iterations it may be observed how each variable becomes increasingly attuned to a particular extension/retraction range (see Fig. 7, *Top*). However, as indicated by Fig. 7, *Middle*, it is difficult to see a pattern or correlation between this attunement and the resulting C_{50} values, which seemingly seem random. Nevertheless, as indicated by Fig. 7, *Bottom*, over time the sum average of the C_{50} values tend to increase, indicating that while individual *genomes* do not evidence positive evolution at each generation, the average of the collective does show important progress over time.

The physical modules shown in Fig. 5 instantiated highest ranked module extension/retraction configurations as expected, only instantiating a new configuration whenever a higher C_{50} value is found or when the user moves to another specific location in the real world as accounted for in the virtual space.

5 Conclusions

In this paper, the development of an acoustically adaptive modular system capable of enhancing Speech Clarity (C_{50} Clarity Index) in specific locations within a space in near real-time is presented. As a CPS, the system consists of a physical/mechanical and a computational component. The form and geometry of the modules conforming the mechanical component enable them (1) to change in the steepness of the sides of their frustum, which changes the way incoming sound waves are deflected/reflected/diffused by the surfaces of the pyramid; and (2) to reveal or to hide the absorbent material under each module, which enables a portion of incoming sound waves to be absorbed/dissipated in a controlled manner. The behavior of these modules is determined by two steps in the computational component of the system. First, the initial position of the modules is set via a model previously generated by an evolutionary solver [15], which identifies the optimal extension/retraction extent of each of the six modules to select for individual configurations that collectively ascertain the highest C_{50} value in said specific locations. Second, a simulated receiver at the location in question measures the actual clarity attained and updates the model's database with respect to the configuration's corresponding clarity-value. The system is conceived as a sub-system in an open-ended and on-going development of an intelligent built-environment framework, and as such it is conceived as one of many solutions designed to enhance well-being and quality of life in the context of AmI/AAL as well as *Interactive Architecture/Architectural Robotics* to instantiate environments that enable was Oosterhuis has called a *Society of Home, Society of Products, Society of Building Components* [21].

Although the system performed as expected, there are several limitations to consider, two of which are the most salient: (1) Sect. 2 mentions that the computational component of the system improves over time. This is true if and only if the selection process is not stuck in *local optima*—that is, a localized optimal value that is only said to be optimal with respect to a particular region of the *landscape* or *volume*. (2) Sect. 3.2 mentions that the system may be said to be evolving positively if subsequent generations are *fitter* than previous ones. This is indeed true, but the danger is that evolutionary solvers do not guarantee it. There may be instances where solutions are stuck in *local optima* and no further progress or positive evolution is observed after a certain iteration. This may be avoided by establishing or configuring mating and mutating-tolerances before running the evolutionary solver. However, it is not possible to eradicate the risk entirely. This makes a strong case for the generation of several models before committing to one. At present, work is being undertaken to implement the system via parallel computational models capable of not only updating within themselves but also across, thereby minimizing the risk of getting stuck in *local optima* by enabling the system to jump from one model to another and to select optimal C_{50} values from all available options.

Acknowledgements. The authors wish to acknowledge Cristian Amaguaña, Juan Balseca, and Dario Cabascango, students of the *Faculty of Electrical & Electronic Engineering at Escuela Politécnica Nacional*, for their assistance in the assembly of the physical implementation. Part of

the present implementation was made possible by funding from *Universidad Internacional SEK's* Project No. P111819.



References

1. Kameas, A., Stathis, K. (eds.): *AmI 2018*. LNCS, vol. 11249. Springer, Cham (2018). <https://doi.org/10.1007/978-3-030-03062-9>
2. De Paz, J.F., Julián, V., Villarrubia, G., Marreiros, G., Novais, P. (eds.): *ISAmI 2017*. AISC, vol. 615. Springer, Cham (2017). <https://doi.org/10.1007/978-3-319-61118-1>
3. Calvaresi, D., Cesarini, D., Sernani, P., Marinoni, M., Dragoni, A.F., Sturm, A.: Exploring the ambient assisted living domain. A systematic review. *J. Ambient Intell. Hum. Comput.* **8**, 239–257 (2017)
4. Grzegorzec, M., Gertych, A., Aumayr, G., Piętka, E.: Trends in Active and Assisted Living - Open hardware architecture, Human Data Interpretation, intervention and assistance. *Comput. Biol. Med.* **95**, 234–235 (2018)
5. Flórez-Revuelta, F., Chaaraoui, A.A. (eds.): *Active and Assisted Living: Technologies and Applications*. The Institution of Engineering and Technology, Stevenage (2016)
6. Fox, M.: *Interactive Architecture: Adaptive World*. Princeton Architectural Press, New York (2016)
7. Green, K.E.: *Architectural Robotics. Ecosystems of Bits, Bytes, and Biology*. The MIT Press, Cambridge (2016)
8. Edelstein, E.A., Macagno, E.: Form follows function: bridging neuroscience and architecture. In: Rasia, S.T., Pardalos, P.M. (eds.) *Sustainable Environmental Design in Architecture. Impacts on Health*, pp. 27–42. Springer, New York (2012). https://doi.org/10.1007/978-1-4419-0745-5_3
9. Roulet, C.-A., Bluysen, P.M., Müller, B., de Oliveira Fernandes, E.: Design of healthy, comfortable, and energy-efficient buildings. In: Rasia, S.T., Pardalos, P.M. (eds.) *Sustainable Environmental Design in Architecture. Impacts on Health*, vol. 56, pp. 83–108. Springer, New York (2012). https://doi.org/10.1007/978-1-4419-0745-5_6
10. Bluysen, P.M.: *The Healthy Indoor Environment. How to Assess Occupants' Wellbeing in Buildings*. Routledge/Taylor & Francis Group, London (2014)
11. Puglisi, G.E., Prato, A., Sacco, T., Astolfi, A.: Influence of classroom acoustics on the reading speed. A case study on Italian second-graders. *J. Acoust. Soc. Am.* **144**, EL144 (2018)
12. Shimizu, T., Onaga, H.: Study on acoustic improvements by sound-absorbing panels and acoustical quality assessment of teleconference systems. *Appl. Acoust.* **139**, 101–112 (2018)
13. Serpanos, D.: The cyber-physical systems revolution. *Computer* **51**, 70–73 (2018)
14. Ochoa, S., Fortino, G., Di Fatta, G.: Cyber-physical systems, internet of things and big data. *Future Gen. Comput. Syst.* **75**, 82–84 (2017)
15. Rutten, D.: Galapagos. On the logic and limitations of generic solvers. *Archit. Des.* **83**, 132–135 (2013)
16. van der Harten, A.: Pachyderm acoustical simulation towards open-source sound analysis. *Archit. Des.* **83**, 138–139 (2013)
17. Grasshopper®: About Grasshopper. <http://www.grasshopper3d.com/>
18. Rutten, D.: *Evolutionary Principles Applied to Problem Solving*. Austria, Vienna (2010)

19. European Association of Research and Technology Organisations (EARTO): The TRL Scale as a Research & Innovation Policy TOOL. EARTO Recommendations. http://www.earto.eu/fileadmin/content/03_Publications/The_TRL_Scale_as_a_R_I_Policy_Tool_-_EARTO_Recommendations_-_Final.pdf
20. Bier, H., Liu Cheng, A., Mostafavi, S., Anton, A., Bodea, S.: Robotic building as integration of design-to-robotic-production and -operation. In: Bier, H. (ed.) *Robotic Building*, 1. Springer International Publishing AG (2018). https://doi.org/10.1007/978-3-319-70866-9_5
21. Oosterhuis, K.: Caught in the act. In: Kretzer, M., Hovestadt, L. (eds.) *ALIVE. Advancements in Adaptive Architecture*, vol. 8, pp. 114–119. Birkhäuser, Basel/Berlin/Boston (2014)



Experiences from Using LoRa and IEEE 802.15.4 for IoT-Enabled Classrooms

Lidia Pocero, Stelios Tsampas, Georgios Mylonas^(✉) ,
and Dimitrios Amaxilatis 

Computer Technology Institute and Press “Diophantus”, Rio, Patras, Greece
{pocero,tsampas,mylonasg,amaxilat}@cti.gr

Abstract. Several networking technologies targeting the IoT application space currently compete within the smart city domain, both in outdoor and indoor deployments. However, up till now, there is no clear winner, and results from real-world deployments have only recently started to surface. In this paper, we present a comparative study of 2 popular IoT networking technologies, LoRa and IEEE 802.15.4, within the context of a research-oriented IoT deployment inside school buildings in Europe, targeting energy efficiency in education. We evaluate the actual performance of these two technologies in real-world settings, presenting a comparative study on the effect of parameters like the built environment, network quality, or data rate. Our results indicate that both technologies have their advantages, and while in certain cases both are perfectly adequate, in our use case LoRa exhibits a more robust behavior. Moreover, LoRa’s characteristics make it a very good choice for indoor IoT deployments such as in educational buildings, and especially in cases where there are low bandwidth requirements.

Keywords: IoT · LoRa · IEEE 802.15.4 · Educational buildings · Real-world deployment · LPWAN · Evaluation

1 Introduction

The smart cities and the Internet of Things (IoT) domains are currently among the most active research areas, having gradually progressed from being mere buzzwords to having actual large-scale installations deployed and applications developed. In this context, a number of competing wireless networking technologies have surfaced in recent years, aiming to appeal to the communities that engage within these two domains. Advancements in wireless communications technology have enabled a multitude of different approaches to the trade-off between power consumption, communication range and bandwidth, in order to answer to all the various types of application use-case requirements. In this context, recent technologies like LoRaWAN [11] and NB-IoT [7] have surfaced, aiming to claim a place in the area originally covered by technologies like ZigBee [12].

L. Pocero and S. Tsampas have contributed equally to this publication.

In the context of the Green Awareness In Action (GAIA) project [2], an IoT platform that combines sensing, web-based tools and gamification elements was developed to address the educational community. Its aim is to increase awareness about energy consumption and sustainability, based on real-world sensor data produced by the school buildings where students and teachers live and work, while also lead towards behavior change in terms of energy efficiency. This real-world IoT deployment provides real-time monitoring of 25 school buildings spread in 3 European countries (Greece, Italy and Sweden).

Due to the multi-year span of the project, a number of conditions, like limited availability of certain networking components and appearance on the market of new ones, have led us to follow a heterogeneous approach with several networking technologies utilized in different buildings of our deployments. During the previous development phases, we have used almost exclusively IEEE 802.15.4-based 2.4 GHz modules. However, school buildings have certain characteristics that in practice lead to less than optimal results in terms of reliability and connectivity. For this reason, we decided to shift towards LoRa-based modules for our deployments in some specific school buildings. LoRa is also well-suited to application use-cases where devices mostly transmit data to the cloud or a nearby gateway (uplink), versus downlink, which also reflects better in the design of other, higher-level, protocols used in IoT. ZigBee and other technologies are better suited for use-cases with more symmetric bandwidth requirements. The frequencies used in LoRa aim for longer range, while they also help to provide a higher degree of wall penetration than other protocols, although 802.15.4 modules are also available in similar frequencies but their availability is limited.

In this paper, we present a comparison between LoRa and IEEE 802.15.4 as the backbone of an IoT deployment inside a number of school buildings. We relay our experiences from using both technologies to develop real-world, reliable and well-performing IoT deployments as a foundation for pervasive computing applications. We present an overview of the two technologies and how we used them, along with an analysis of the effect of changing parameters like network density, application data rate and distance between nodes. Our results indicate that in our use-case and under the design constraints that we had, LoRa works in a more reliable manner while also satisfying our data rate requirements.

2 Previous Work

Regarding recent comparisons between protocols used for low power wide area networks (LPWAN) in IoT, [21] and [25] discuss aspects related to LoRa, NB-IoT and ZigBee. All of these technologies are being used especially in smart city applications, and currently there is a lot of interest in understanding the parameters related to their performance in the real world. This aspect is discussed in [22], where a smart city deployment using LoRa and IEEE 802.15.4 is evaluated using mostly simulation methods and limited real-world studies. [17] provides a survey of LoRaWAN for IoT and recent examples of related applications, along with a discussion on its advantages and shortcomings.

Although most related works describing aspects like the ones mentioned above are limited to simulation, some recent ones performing measurements in real-world settings. LoRa performance is explored to a certain degree in [15], with a discussion on possibilities and limitations. In [14], a performance evaluation of LoRaWAN and its integration in IoT devices is discussed, while [19] explores its scalability in the context of large-Scale sensor networks. [20] presents an evaluation of LoRaWAN using a permanent outdoor deployment, while [23] provides an experimental study on LPWANs for mobile IoT applications. [16] provided a study of LoRa in long-range use-cases and produced certain radio propagation models to be used when designing LoRa-based solutions. Their work confirmed coverage of up to 8Km in urban and 45Km in rural areas (with line-of-sight). [25] provided a simulation-based comparative study between LoRa and NB-IoT, describing the advantages of each technology in specific areas and use-cases.

However, so far work is either mostly based on simulation, or they do not attempt a straight apples-to-apples comparison between different networking technologies in specific use-cases e.g., for IoT, pervasive computing or smart cities. Our work here contributes to the discussion over which technology is better suited for real-world application in a representative use-case; school buildings are a characteristic and ubiquitous example of public building. Our application requirements in terms of data sampling and quality of service (QoS) are also similar to other related application scenarios (e.g., office building monitoring and automation). While there are a number of works presenting open source power meter similar to the ones used in GAIA, they do not focus much on wireless communication and protocol aspects. Regarding our own previous work, the design and implementation of the cloud-based aspects of the project are discussed in [13], while open source hardware aspects are discussed in [24].

3 Short Overview of IEEE 802.15.4 and LoRa

In this section, we present a brief comparison between the IEEE 802.15.4 and LoRa networking, in order to give a context for the sections that follow and discuss their performance in more detail. Overall, although the 2 protocols overlap in their use in real-life applications, they have different design goals and, at least in theory, target different domains. LoRa is the more recent technology of the two, targeting (as its name suggests) long-range use-cases, while 802.15.4 targets more conventional applications. However, in practice the two technologies currently are used in similar contexts in many cases, and this is the reason why we compare them in this paper: we wish to examine their respective advantages and limits in a typical IoT infrastructure scenario inside a building.

3.1 IEEE 802.15.4

The IEEE 802.15.4 protocol specification includes both a Physical and a MAC layer definition. The physical layer defined the frequency (possible frequencies are 868 MHz, 915 MHz and 2.4 GHz) and the number of channels. The MAC layer

defines the device types (physical address) and channel access. The 802.15.4 physical layer defines the possibility of 16 channels in ISM band from 5 MHz channel spacing, beginning at 2405 MHz and ending at 2480 MHz. The carrier-sense multiple access with collision avoidance (CSMA/CA) protocol is implemented as part of the MAC layer by using a CCA (clear channel assessment) technique to determine if the channel is available before transmitting a packet [6]. The detected energy on the channel is compared with the CA (Clear Channel Assessment) parameter value. If the detected energy exceeds the CA parameter value, the packet is not transmitted. The module will attempt to re-send the packet up to two additional times, in case of CCA failure.

Moreover, the European Telecommunication Standards Institute (ETSI) regulates the maximum transmitted RF power in wireless networking modules via the ETSI EN 300 328 standard [1]. Two clauses are the most important: the maximum transmit power, limiting power to 100 mW, and the maximum EIRP spectral density, which is limited to 10 mW/Hz [6]. The ETSI standard sets a safe limit for RF output power around 12 dBm [1]. Furthermore, in the 2.4 GHz band, a maximum over-the-air data rate of 250 kbps is specified, but due to the overhead of the protocol, the actual maximum data rate is approximately half [3].

For our network implementation, for the 802.15.4 part we have chosen to use XBee network modules; in the rest of the text, XBee refers to 802.15.4. We set every XBee module at the 802.15.4 MAC mode with ACKs acknowledgment protocol. The RF module operates in a unicast mode that supports retries. The receiving modules send an ACK of RF packets to confirm reception to the transmitter. If the transmitting module does not receive the ACK, it will resend the packet up to three times, or until the ACK packet is received. The transmission happens directly without any delays. The modules are configured to operate with a peer-to-peer network topology with no master/slave relationship and each module of the network shares both roles master and slave. The Network ID and Channel must be identical across all the modules in the network. Each RF packet contains a maximum of 100 characters (100 bytes). In our network, the payload of the RF packet will be variable but always smaller than the 100 character limit, which means all messages are transmitted within one packet.

3.2 LoRa

Long Range (LoRa) was originally conceived as a long-range wireless communication technology that operates on the sub-GHz license free ISM bands (868 MHz in Europe and 915 MHz in the U.S.). This means that, in contrast to other related technologies like NB-IoT, it operates in frequencies that are free to use and anybody can potentially operate a LoRa network without requiring a license for it. Regarding features of LoRa that are examined in this work, the over-the-air LoRa modulation technique can be understood as a MFSK modulation on top of a Chirp Spread Spectrum (CSS) method. Each bit is spread by a chipping factor, with the number of chips per bit called Spread Factor (SF). Chirps are used to encode data in LoRa networks for transmission, while inverse chirps are

used on the receiver side for signal decoding. The modulation across the channel is sweeping so that the transmission signal occupies the chosen bandwidth (BW). SFs specifically set the data transfer rate relative to the range, by essentially indicating how many chirps are used per second, and define bit rates, per symbol radiated power, and achievable range. The possible values of SF are between 6 and 12. The data rate depends on the selected SF, e.g., *SF9* is 4 times slower than *SF7* in terms of bit rate. In general, the slower the bit rate, the higher the energy per data set and the higher the range [14].

The 868 ISM frequency band ranges from 865 MHz to 870 MHz and is regulated for the European zone [5]. The rules are based on two restrictions: (a) the maximum power transmission that can be used on a channel at the communication is 25 mW (equivalent of 14 dB); (b) the duty cycle that is defined as the ratio of maximum time-on-air (ToA) per hour and is limited to 1%, which in practice restricts the communication of each LoRa device with other nodes to 36 s per hour [18]. The MAC layer of LoRa does not implement any listen-before-talk (LBT) or CSMA to avoid collisions. Instead it implements a pure Aloha protocol, sending data whenever available, thus the number of collisions increases together with transmission rate or network node density.

4 A Large-Scale IoT Infrastructure Inside School Buildings

Our IoT-based platform aims to provide an integrated solution for real-time monitoring and management of a large fleet of educational buildings with diverse characteristics. The platform's goal is to build upon IoT networks that operate in a vast number of buildings and spread over several rooms inside each school buildings, while incorporating various kinds of sensors. Given the diverse building characteristics and usage requirements, deployments vary between schools (e.g., number of sensors, manufacturer, networking, etc.). The deployed devices provide 1250 sensing points organized in four categories: (1) classroom environmental sensors; (2) atmospheric sensors (outdoors); (3) weather stations (on rooftops); and (4) power consumption meters (attached to electricity distribution panels). These devices are connected to cloud services via IoT gateway devices, which coordinate communication with the rest of the platform, while outdoor nodes use wired networking or WiFi. The IoT devices (Fig. 1) used are either open-design IoT nodes, or off-the-shelf products from IoT device manufacturers. Indoor devices use IEEE 802.15.4 in a tree-routing topology or LoRa wireless networks in a star network (Fig. 2). In this paper we compare the performance of these two IoT networking technologies in our real world implementations.

4.1 IEEE 802.15.4 Network Topology

The IEEE 802.15.4 communication between the IoT nodes is provided by XBee modules connected to each IoT node operated by the Arduino XBee [10] and XBeeRadio [9] software libraries. Node-to-node communication includes the

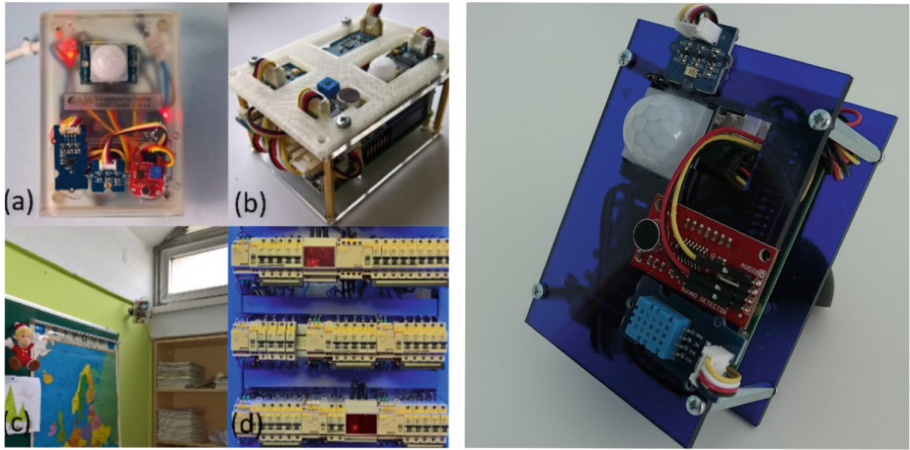


Fig. 1. Examples of the IoT infrastructure located inside school buildings in Greece (a–b) IoT nodes based on Arduino and Raspberry Pi, (c) actual node inside a classroom, (d) power meter installed on a electricity distribution panel. On the right part of the figure, the latest hardware revision of the actual environmental nodes used in our IoT infrastructure, utilizing a LoRa communication module.

checksum of the payload, which is validated at the network level for each node to determine erroneous or invalid messages, which are discarded.

For the network layer, all IoT nodes form ad hoc networks and report their measurements through the designated IoT gateways. Because IEEE 802.15.4 is a short-range communication technology and end-to-end communication is not possible due to power limitations and propagation obstacles, all indoor IoT nodes form an ad hoc overlaying multi-hop bidirectional tree network. The gateway is the root of this tree and the orchestrator of the network. On boot and at regular intervals the gateway node broadcasts heartbeat messages. Upon receiving a heartbeat, non-gateway nodes decide to join the tree and re-broadcast the heartbeat. The re-broadcasts occur to propagate the heartbeat of the gateway to the nodes that do not have direct connectivity with it. New nodes can join the network at any time either directly below the gateway, or as a child of the node that is closer to the gateway and has a received signal strength indication (RSSI) lower than a specific threshold (in our case 90 dB). The resulting routing tree allows for bidirectional communication between the IoT nodes and the gateway. The routing library developed for the Arduino and XBee devices is also available on GitHub [7]. An example of a formed network can be seen in Fig. 2(left).

Once the network has been established, each node collects environmental or other sensor data and emits a data packer (e.g., an *Environmental Data Packet (EDP)*) to the GW every 10s. The payload size varies depending on the sensing activity, but in our case it is always lower than the limit of a 100 character payload to fit in a single packet. In addition, each environmental node

checks its motion (PIR) sensor every 2s and emits a *PIR Data Packet (PDP)* independently each time motion is detected.

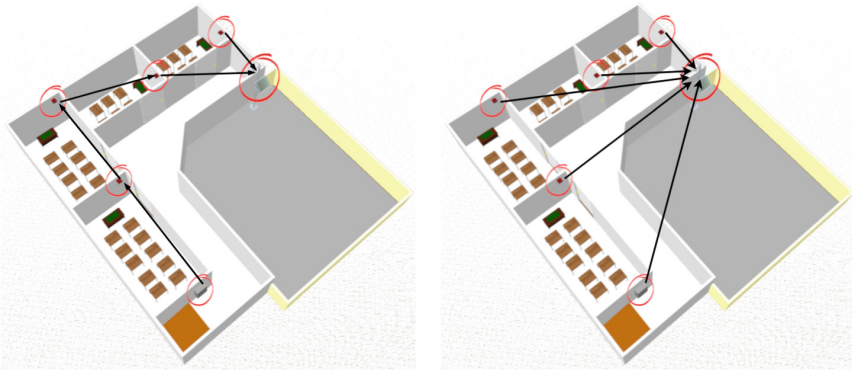


Fig. 2. Examples of data collection routes in an IEEE 802.15.4 (left) and a LoRa (right) IoT deployment, using a tree/multihop and a star network topology, respectively.

4.2 LoRa Network Topology

Our LoRa nodes use a single-hop topology to cover the necessary distance (tested within 3-floor concrete-built buildings), thanks to their communication range and signal penetration. An example of a formed network showing the difference with the IEEE 802.15.4 network can be seen in Fig. 2(right). We use LoRa to build our own wireless LoRa Personal Area Network (PAN) with a star topology. The IoT nodes communicate using the Grove LoRa 868 MHz [3] modules with a LG01-N Single Channel LoRa Dragino gateway [8], suitable for small-scale LoRa networks. The communication device for both the Dragino and the Grove modules is based on the RF95 SX1276 LoRa module [4]. The GW coordinates the communication with every node to guarantee that the nodes do not occupy the medium at the same time. This implementation is necessary to avoid the interference due to ALOHA MAC protocol and guarantee no interference between the nodes at the same network. The network is created by the GW announcing itself through broadcast messages. The new nodes reply to the broadcast with a connection request which, if it is accepted by the GW, is acknowledged by a confirmation message. The ALOHA protocol with a randomized delay is used by the nodes to answer to the GW network announcement.

Once the network is setup, the GW requests data periodically from each node in a sequential fashion with a *Data Request Packet (DRP)*. The nodes reply with a *Data Packet (DP)* consisting of the sensor measurements. The request rate of the GW is configurable to adjust to the requirements of ToA EU regulations. In this case, the node samples the PIR sensor between GW requests and includes

the motion sensing information in the DP to avoid creating overhead. If a reply is not received or the reply is corrupted, the GW can repeat the DRP up to three times for each node. We implemented our own *CRC (Cycle Redundancy Check)* method at network level to detect message corruption instead of using the LoRa module functionality at MAC level. The network is refreshed every 15 min. On each refresh, new nodes can be attached while unreachable nodes are not removed to speedup future reconnects.

5 LoRa Network Configurations Comparison

The maximum *ToA* is restricted in Europe, thus limiting the packet rate for each network device. The *ToA* of each LoRa packet depends on the *spreading factor* (SF), *coding rate* (CR), *signal bandwidth* (BW) and the *packet payload* (PL). The LoRa packet duration is the sum of the duration of the preamble and the transmitted packet. The data-sheet of the SX1276 module [4] describes the formula to calculate the number of payload symbols and the preamble length. We use this to determine the *ToA* of each packet of our network in milliseconds, and thus we can calculate the maximum legal packet rate to accommodate the 36 s *ToA* per node limit under different network configurations. The maximum *PL* of each *DP* is 60 bytes and the *PL* of each *DRP* is fixed at 4 bytes. A comparison for each type of packet between different network configuration is described in Table 1 showing the corresponding *PL*, *ToA* for a single packet in milliseconds and the minimum *Period* between transmissions in seconds.

Table 1. Packet *ToA* and Period per device

SF	BW [kHz]	Node (data packets)			Gateway (data request packets)		
		PL [Bytes]	Packet <i>ToA</i> [ms]	Min. period [s]	PL [Bytes]	Packet <i>ToA</i> [ms]	Min. period [s]
7	125	60	112.896	11.289	4	30.976	3.097
7	250	60	56.448	5.644	4	15.488	1.548
7	500	60	25.088	2.8224	4	7.744	0.774
9	125	60	319.488	36.966	4	123.904	12.390
9	250	60	159.744	18.483	4	61.952	6.195
9	500	60	79.872	9.241	4	30.976	3.097

As described, the GW requests data from each node periodically by sending a *DRP* with a 4 bytes payload. As a consequence, in our case the maximum packet rate (minimum period between *DPs*) per node is limited by the maximum packet rate (minimum period between *DRPs*) of the GW, which depends on the total number of nodes in the network. Table 2 presents the theoretical minimum *period* per node and the maximum *packets* per 15 min in our network as influenced by

Table 2. Theoretical minimum Period and maximum Packets per 15 min. n represents the number of nodes in the network

SF	BW [kHz]	n Nodes		LoRa School A (6 nodes)		LoRa School B (7 nodes)	
		Minimum period [s]	Maximum packets[#]	Minimum period[s]	Maximum packets[#]	Minimum period[s]	Maximum packets[#]
7	125	$3.09*n$	$290.54/n$	18.58	48.42	21.68	41.50
7	250	$1.54*n$	$581.09/n$	9.29	96.84	10.84	83.01
7	500	$0.77*n$	$1162.19/n$	4.64	193.69	5.42	166.02
9	125	$12.39*n$	$72.63/n$	74.34	12.10	86.73	10.37
9	250	$6.19*n$	$145.27/n$	37.17	24.21	43.36	20.75
9	500	$3.09*n$	$290.54/n$	18.58	48.42	21.68	41.50

the restrictions of the *ToA* of the GW under different network configurations for two schools (*LoRa School A*, *LoRa School B*) of our installation.

The maximum *packet rate* per node is achieved with *SF* 7 and *BW* 500 kHz, which we implemented in our final network installation due to our priority of maximizing the sensing rate in the school and achieving a better sampling of the environmental reality in the public buildings. It is noteworthy that higher spreading factors allows for longer range at the expense of lower data rate, and vice versa. We aim to compare the quality of the network under two extreme configurations. Configuration A provides higher rate (*SF* 7, *BW* 500 kHz) and Configuration B provides longer range (*SF* 9, *BW* 125 kHz). In order to study the network behaviour, we collected the following measurements per node: the number of *DRPs* from the GW, the number of received *DPS*, and the number of packets received with *CRC* errors over a period of 15 min.

The maximum number of packets received under Config. 1 is limited by the *ToA* imposed by communications regulations. Thus, we have to set the GW request rate accordingly to implement this restriction. As such, Configuration 1 is limited to 12 packets, per 15 min, per node. On the other hand, Config. 2 is limited by the regulation at 193.69 (Table 2) packets, per 15 min, per node. Effectively, Configuration 2 is restricted by the node design constraints. The GW requests data from each node after a 50 ms delay to guarantee the correct communication between the LoRa module and the micro-controller. In addition, the Environmental Nodes consume time to communicate through I2C with their digital sensors to collect the data for each request, limiting the final rate of the node. Due to these factors, in School Building A’s installation every node can achieve a maximum of 174 packets (Table 3) per 15 min, which is lower but close to the theoretical limit. On both configurations, the average of the delivered *DP* rate (Table 3) is higher for the nodes near the GW (1, 5, 6).

As an indicator of the quality of the network, we use *CRC Error Ratio* and *Re-transmission Ratio*. *CRC Error Ratio* is the ratio of *DRPs* from the GW which resulted in a corrupted *DP* being received. *Re-transmission Ratio* is the ratio of *DRPs* required to be repeated, either because of *CRC* errors, malformed *DRP* or due to not receiving a reply. We are also interested in the connectivity

Table 3. Number of delivered *DPs* per node under different configurations in a 15 min period

		node1	node2	node3	node4	node5	node6
Configuration 1 SF 9, BW 125 kHz	Avg.	11.17	11.17	11.15	11.16	11.16	11.16
	Min.	5.00	5.00	0.00	0.00	5.00	5.00
	Max.	12.00	12.00	12.00	12.00	12.00	12.00
Configuration 2 SF 7, BW 500 kHz	Avg.	155.24	149.77	145.79	150.87	155.23	155.22
	Min.	77.00	0.00	0.00	1.00	77.00	77.00
	Max.	174.00	174.00	174.00	174.00	174.00	174.00

between the GW and every node in our LoRa network. We quantify the quality of each link by calculating the *Packet Delivery Ratio (PDR)* for every node (Table 4). The PDR of the link between node A and GW can be measured as the ratio between the number of *DPs* received by the GW from node A, and the number of *DRPs* sent from the GW to node A. The GW makes one *DRP* and a maximum of 3 *DRP* re-transmissions per node. In addition, we study the variation of the *RSSI* per node in the network for both configurations (Table 5).

Table 4. Packet Delivery Ratio (PDR) per node under different configurations

		node1	node2	node3	node4	node5	node6
Configuration 1 SF 9, BW 125 kHz	Avg.	0.96	0.96	0.96	0.95	0.95	0.93
	SD	0.03	0.03	0.04	0.05	0.03	0.03
	Min.	0.72	0.80	0.00	0.00	0.86	0.80
	Max.	1.00	1.00	1.00	1.00	1.00	1.00
Configuration 2 SF 7, BW 500 kHz	Avg.	0.99	0.93	0.86	0.89	0.98	0.97
	SD	0.01	0.22	0.27	0.19	0.01	0.02
	Min.	0.91	0.00	0.00	0.00	0.93	0.88
	Max.	1.00	1.00	1.00	1.00	1.00	1.00

We expected to observe a worse quality network under Config. 2, as a consequence of selecting parameter values that achieve a higher packet rate. We can observe that the median value for the *CRC Error Ratio* distribution of each node is higher with similar standard deviation with the exception of the closest node (Node 1, *CRC Error Ratio* graph in Fig. 3). The degradation of network quality is evident at the farthest nodes (3 and 4) from the GW regarding the number of Re-transmissions (Fig. 3), which exhibits higher standard deviation and more frequent and distant upper outliers. In addition, the PDR (Fig. 3) of these nodes is worse than in Configuration 1, with higher standard deviation and more frequent and distant lower outliers. On the other hand, the nodes closest

Table 5. RSSI per node under different configurations

		node1	node2	node3	node4	node5	node6
Configuration 1 Sf 9, BW 125 kHz	Avg.	-45.06	-50.55	-87.78	-87.08	-53.52	-52.40
	SD	0.41	0.98	2.59	1.73	1.35	1.29
	Min.	-46.13	-54.25	-95.47	-95.29	-59.40	-55.77
	Max.	-43.74	-48.75	-82.12	-84.12	-51.73	-50.17
Configuration 2 Sf 7, BW 500 kHz	Avg.	-42.04	-46.29	-82.23	-86.86	-49.88	-44.78
	SD	3.17	10.84	16.92	4.66	1.68	3.25
	Min.	-55.33	-56.98	-89.00	-88.44	-57.70	-60.79
	Max.	-37.40	0.00	0.00	0.00	-48.17	-41.27

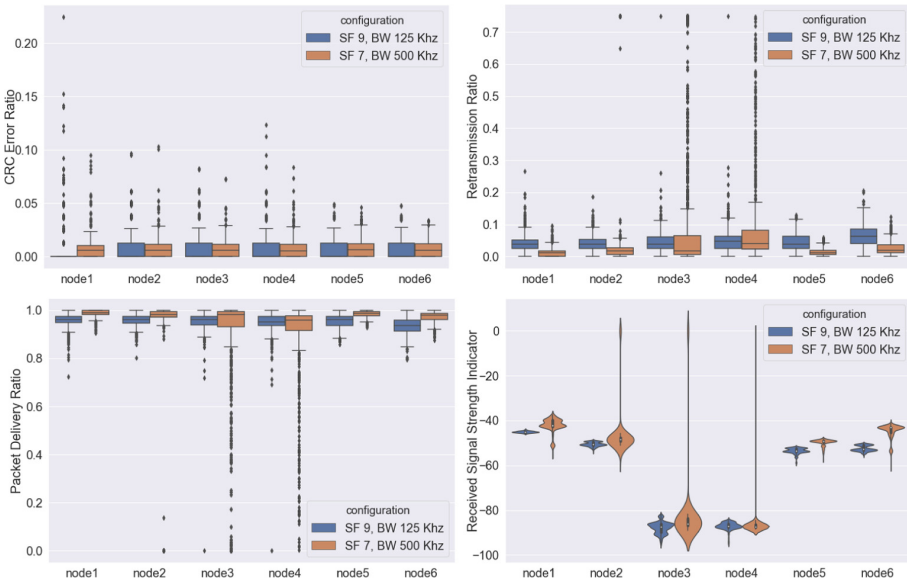


Fig. 3. Per node statistics for configurations 1: *SF 7, BW 500kHz* and 2: *SF 9, BW 125kHz*.

to the GW achieve better network behavior regarding Packet Delivery and Retransmission Ratios. Furthermore, the RSSI distribution (Fig. 3) exhibits greater standard deviation, entailing less stable signal strength.

In conclusion, we observed the expected cost in network quality, only in the further nodes, while in nearby nodes we observed an increase in the link's efficiency. This combined with the increase in the per node packet rate, resulted in a significant increase in sensor measurements across the whole network.

6 XBee Network Behavior in School Buildings

Every node in the XBee network tries to deliver an *Environmental Data Packet (EDP)* to the GW every 10s while emitting an extra *PIR Data Packet (PDP)* each time motion is detected. In our XBee network the CSMA/CA protocol is used, which provides a limited communication reliability, due to the slotted channel access algorithm. When a large number of nodes try to transmit simultaneously, it is less probable to find empty slots to access the channel. The extra *PDPs* generated can saturate the channel and provoke a decrease of *EDPs* delivered per node, thus decreasing the EDP rate. The data-set considered to analyze the specific behaviour in XBee School C is composed of the total number of packets delivered in the network in 5 min periods (*EDPs* and *PDPs* from every nodes). To quantify the effects of the independent *PDPs*, we use their ratio against the observed maximum of the aggregation.

Figure 4 shows clearly how during school hours the *PDPs* can cause an observable decrease in the number of *EDPs*, due to the saturation of the network at peak of *PDP* Ratio. The number of *PDPs* exhibits a maximum, because of the peak of activities in the school building, the number of *EDPs* decreases below their average. In addition, when the maximum number of *EDPs* is observed the number of *PDPs* is zero (Table 6), effectively when there is no movement detected. The decision to include real-time motion detection to the network, can potentially be a hindrance to the stability of the *EDP* rate of our network.

Table 6. XBee Network behaviour. Number of *Aggregated Packets*, *PDPs* and *EDPs* delivered at the time of maximum and minimum *Aggregated Packet Ratio*, *EDP Ratio* and *PDP Ratio* respectively. The vertical lines indicate the maximum of each ratio in the corresponding colour.

	Average	Aggregated packets		PIR packets		Env. packets	
		Max	Min	Max	Min	Max	Min
Aggregated packets [#]	143.05	185	84	178	142.11	157	87
PIR packets [#]	1.91	41	3	45	0.00	0	7
Node packets [#]	141.13	144	81	133	142.11	157	80

7 Discussion - Comparison Inside School Buildings

We aim to compare the quality of our LoRa and IEEE 802.15.4 IoT networks by comparing our observations from two real school buildings. LoRa School A consists of a LoRa Network with 6 nodes where the Node 3 and 4 are located at the farthest positions and the Node 1 and 5 at the nearest in relation with the GW. XBee School C has an XBee network consisting of 6 nodes, the farthest is node 6 and the nearest is the node 5. Due to the significant differences in network architectures, to compare them we quantify the quality of these networks by the

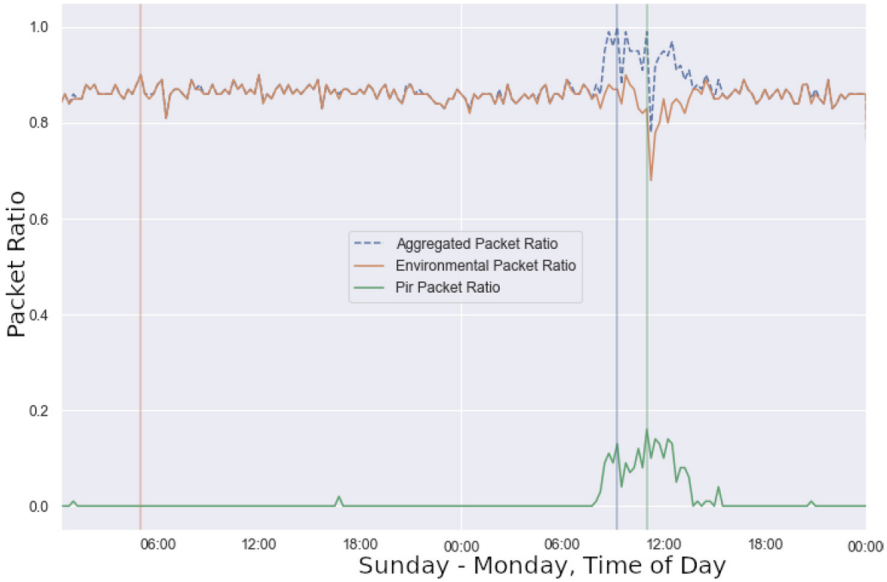


Fig. 4. *Aggregated Packet Ratio, EDP Ratio and PDP Ratio in XBee School C over 48 h (Sunday and Monday)*

Network Delivery Ratio (NDR). The *NDR* is defined as the ratio between the measured Delivered Packets and the potential maximum number of Packets that could be delivered by each node in the network in a time period, which in our case is 15 min. In effect, we try to quantify the end-to-end success of packet deliveries from a Node to the GW, disregarding how the transmissions are routed through single or multi hop network topologies and MAC and LAN protocols. Every node in the XBee network is scheduled to attempt to send data to the GW every 10 s, resulting in a maximum of 90 packets in a 15 min period. In the LoRa network using Configuration 2 and with 6 nodes, the network can achieve the delivery of a maximum of 174 packets. The data-set consists of *NDR* measurements collected during a period of both business and weekend days for LoRa School A and XBee School C can be seen in Table 7 and Fig. 5.

We observe that the best network quality in terms of *NDR* is observed in the LoRa School A under Config. 1, which is the one with lowest rate: the maximum number of Data Packets per node in a 15-min. period is 12 packets. The network in XBee School C exhibits a better *NDR* than LoRa School A under Config. 2, regarding the averages for every node with the exception of the farthest one that achieves a delivery of 20% of the generated packets. On the other hand, the network in LoRa School A under Config. 2 achieves a more stable *NDR* across the installation, including the farthest nodes, with successful packet deliveries between 86% and 89% for every node and a significantly higher delivery rate. In order to achieve packet rate equivalent to XBee School C, it is required to use Config. 2 in LoRa School A. Although the network quality of Configuration

2 fluctuates, as evident by its significant standard deviation, it achieves better *NDR* in distant devices (Table 7 node 6). The tree topology, necessary for XBee to achieve comparable range to LoRa, negatively influences the packet rate of

Table 7. Network Delivery Ratio (*NDR*) per node for LoRa and IEEE 802.15.4 Networks in different school buildings

Network		node1	node2	node3	node4	node5	node6
LoRa School A Conf. 1 SF 9, BW 125	Avg.	0.94	0.94	0.94	0.94	0.94	0.94
	SD	0.03	0.03	0.04	0.04	0.03	0.03
	Min	0.41	0.41	0.00	0.00	0.41	0.41
	Max	1.00	1.00	1.00	1.00	1.00	1.00
LoRa School A Conf. 2 SF 7, BW 500	Avg.	0.89	0.86	0.84	0.87	0.89	0.89
	SD	0.12	0.22	0.25	0.17	0.12	0.12
	Min	0.44	0.00	0.00	0.01	0.44	0.44
	Max	1.00	1.00	1.00	1.00	1.00	1.00
XBee School C	Avg	0.85	0.91	0.92	0.92	0.92	0.20
	SD	0.04	0.05	0.05	0.04	0.05	0.05
	Min	0.28	0.32	0.31	0.31	0.32	0.09
	Max	0.91	0.98	0.98	0.98	0.99	0.38

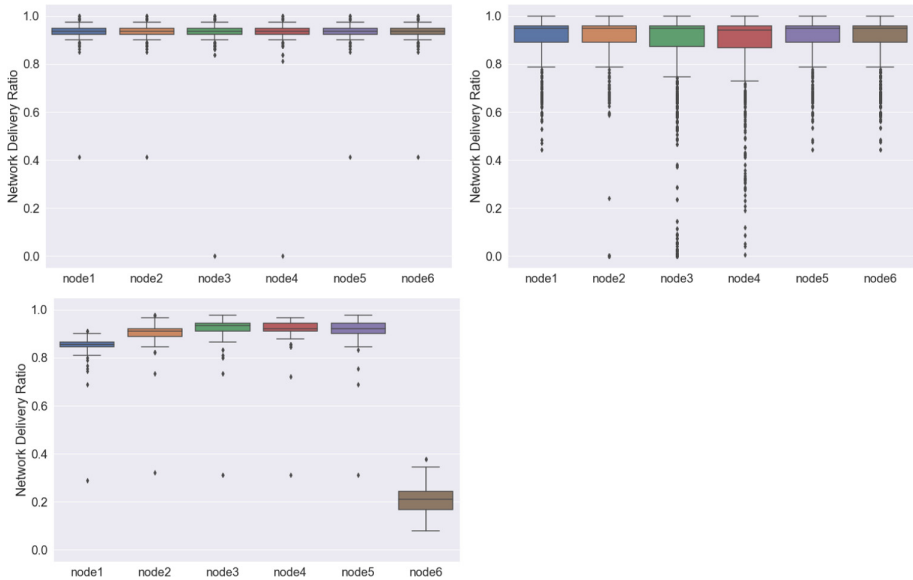


Fig. 5. Network Delivery Ratio (*NDR*) per node in LoRa School A under Configuration 1 (top row, left) and Configuration 2 (top row, right) and XBee School C (bottom row)

the nodes placed at the extremes of the tree. In comparison, LoRa's star network topology offered better coverage with a more stable data rate on all nodes.

8 Conclusions and Future Work

Our work in recent years has resulted in the deployment of a large-scale IoT infrastructure inside a number of school buildings in Europe. In this context, we have opted to use different wireless networking technologies in order to test in practice their performance. With this work, we wanted to relay our practical experiences from using both IEEE 802.15.4 and LoRa for our specific application use-case and provide some practical examples and guidelines for IoT deployments that are similar to ours. We have studied the behavior of both networks, in the scenario of changing the number of nodes in the network, varying the sampling rate of the sensors and the required data rate, or changing the distance between the IoT nodes inside the building. As an example of the results from the comparisons we made, LoRa decreases its delivery rate when increasing the number of nodes because of ToA European regulations which restricts the number of GW data requests in our network design. In 802.15.4 we expect an increased number of collisions when adding nodes due to CSMA. In the case of increasing the distance between nodes, LoRa achieves longer range with a stable rate, while 802.15.4 will need hop nodes in the middle, leading to increased number of collisions and an unstable rate in extreme nodes as a side effect. Overall, our results show that in the use-case scenario and environmental settings of school buildings in Greece, LoRa-based wireless communication can have an advantage against competing technologies, in terms of reliability and complexity of networking. Regarding our future work, we plan to conduct a more thorough performance evaluation and explore in additional dimensions practical aspects like networking performance and reliability.

Acknowledgment. This work has been supported by the EU research project "European Extreme Performing Big Data Stacks" (E2Data), funded by the European Commission under H2020 and contract number 780245, and the "Green Awareness In Action" (GAIA) project, funded by the European Commission and the EASME under H2020 and contract number 696029. This document reflects only the authors' views and the EC and EASME are not responsible for any use that may be made of the information it contains.

References

1. Final draft ETSI EN 300 328 V2.2.1: Wideband transmission systems; Data transmission equipment operating in the 2,4 GHz band; Harmonised Standard for access to radio spectrum. <https://www.etsi.org>. Accessed 19 July 2019
2. GAIA project website. <http://gaia-project.eu/>. Accessed 19 July 2019
3. Grove-LoRa Radio 868 MHz. <https://www.seeedstudio.com/Grove-LoRa-Radio-868MHz-p-2776.html>. Accessed 19 July 2019

4. Semtech, sx1276/77/78/79 datasheet. <https://www.semtech.com/products/wireless-rf/loro-transceivers/sx1276>. Accessed 19 July 2019
5. SX1272/3 LoRa Modem. <https://www.semtech.com/uploads/documents/etsi-compliance-sx1272-lora-modem.pdf>. Accessed 19 July 2019
6. XBee/XBee-PRO S1 802.15.4 (Legacy) RF Modules User Guide. <https://www.digi.com/resources/documentation>. Accessed 19 July 2019
7. Narrowband - Internet of Things (NB-IoT). <https://www.gsma.com/iot/narrowband-internet-of-things-nb-iot/>. Accessed 4 Sept 2019
8. LG01-N Single Channel LoRa IoT Gateway. <https://www.dragino.com/products/lora/item/143-lg01n.html>. Accessed 19 July 2019
9. mksense, arduino xbee radio library. <https://github.com/mksense>. Accessed 19 July 2019
10. Arduino XBee library. <https://github.com/andrewrapp/xbee-arduino>. Accessed July 2019
11. Alliance, L.: LoRaWan specification v1.1, p. 100 (2017)
12. Alliance, Z.: Zigbee specification document 053474r20, p. 622 (2012)
13. Amaxilatis, D., Akriopoulos, O., Mylonas, G., Chatzigiannakis, I.: An IoT-based solution for monitoring a fleet of educational buildings focusing on energy efficiency. *Sensors* **17**(10), 2296 (2017)
14. Bouras, C., Kokkinos, V., Papachristos, N.: Performance evaluation of LoRaWan physical layer integration on IoT devices, pp. 1–4 (2018). <https://doi.org/10.1109/GIIS.2018.8635715>
15. Carlsson, A., Kuzminykh, I., Franksson, R., Liljegren, A.: Measuring a LoRa network: performance, possibilities and limitations. In: Galinina, O., Andreev, S., Balandin, S., Koucheryavy, Y. (eds.) NEW2AN/ruSMART -2018. LNCS, vol. 11118, pp. 116–128. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01168-0_11
16. El Chall, R., Lahoud, S., El Helou, M.: LoRaWan network: radio propagation models and performance evaluation in various environments in Lebanon. *IEEE Internet of Things J.* **PP**, 1 (2019). <https://doi.org/10.1109/JIOT.2019.2906838>
17. Haxhibeqiri, J., De Poorter, E., Moerman, I., Hoebeke, J.: A survey of LoRaWAN for IoT: from technology to application. *Sensors* **18**(11) (2018). <https://doi.org/10.3390/s18113995>
18. Lavric, A., Popa, V.: A LoRaWAN: long range wide area networks study. In: 2017 International Conference on Electromechanical and Power Systems (SIELMEN), pp. 417–420 (2017)
19. Lavric, A., Popa, V.: Performance evaluation of LoRaWan communication scalability in large-scale wireless sensor networks. *Wirel. Commun. Mob. Comput.* **2018**, 1–9 (2018). <https://doi.org/10.1155/2018/6730719>
20. Marais, J., Malekian, R., Abu-Mahfouz, A.: Evaluating the LoRaWan protocol using a permanent outdoor testbed. *IEEE Sens. J.* **PP**, 1 (2019)
21. Mekki, K., Bajic, E., Chaxel, F., Meyer, F.: A comparative study of LPWAN technologies for large-scale IoT deployment. *ICT Express* **5**(1), 1–7 (2019). <https://doi.org/10.1016/j.icte.2017.12.005>
22. Pasolini, G., et al.: Smart city pilot projects using LoRa and IEEE802.15.4 technologies. *Sensors* **18**, 1118 (2018). <https://doi.org/10.3390/s18041118>
23. Patel, D., Won, M.: Experimental study on low power wide area networks (LPWAN) for mobile internet of things. In: 2017 IEEE 85th Vehicular Technology Conference (VTC Spring), pp. 1–5, June 2017. <https://doi.org/10.1109/VTCSpring.2017.8108501>

24. Pocero, L., Amaxilatis, D., Mylonas, G., Chatzigiannakis, I.: Open source IoT meter devices for smart and energy-efficient school buildings. *HardwareX* **1**, 54–67 (2017). <https://doi.org/10.1016/j.ohx.2017.02.002>
25. Sinha, R.S., Wei, Y., Hwang, S.H.: A survey on LPWA technology: LoRa and NB-IoT. *ICT Express* **3**(1), 14–21 (2017). <https://doi.org/10.1016/j.icte.2017.03.004>



Adaptive Service Selection for Enabling the Mobility of Autonomous Vehicles

Elif Eryilmaz¹(✉), Manzoor Ahmed Khan¹, Frank Trollmann²,
and Sahin Albayrak¹

¹ Technische Universität Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany
elif.eryilmaz@dai-labor.de

² CODE University of Applied Sciences, Lohmhlenstrae 65, 12435 Berlin, Germany

Abstract. Recent advances in the Internet of Things (IoT) provides rich opportunities to future mobility services for the development of more flexible solutions. Instead of using a fixed set of data sources or services, applications can benefit from those flexible mechanisms by adapting to change in the sensing environment such as sensor disappearance/degradation or service unavailability. In this paper, we contribute with an approach that enables dynamic selection of the services for mobility to meet requirements of autonomous driving use-cases. Our approach is focused on different mobility services using available data sources and data processing methods with their related quality parameters. Those services are inspired by the standards and the dynamics of real-test road. We present a prototypical implementation of the mechanism for optimal service selection in an autonomous driving test environment and evaluated our testing results with respect to correctness and performance.

Keywords: Adaptive service selection · Quality-optimal selection · Opportunistic sensing · Autonomous driving · Mobility · IoT

1 Introduction

Advancement of sensor technologies, availability of immense compute infrastructure, and milestones recently achieved by the communication technologies (e.g., 5G standardization by 3GPP [1]) may consequence in earlier realization of the autonomous driving vision than expected. It stands to reason that autonomous driving involve various technologies and concepts e.g., fusing the sensory data, creating the vehicle perception, evolved Advanced Driver Assistance Systems (ADAS), etc. A closer look into the operations of envisioned autonomous driving technologies and their inter-play reveals that architectures have evolved from reference point architecture to the service-based architectures. Autonomous driving services are being developed very rapidly by different stakeholders around the world e.g., Original Equipment Manufacturers (OEMs), network operators, sensor providers, road infrastructure provider/operators, etc. Obviously, the maturity level of these services and their utilization in the real-world scenarios vary

and in most cases are undergoing experimental trials. The idea of implanting human perception and cognition into autonomous vehicles rely on multiple fronts including: deploying additional sensors on the vehicle, improving the sensory fusion for different use cases, providing vehicles with additional information from external sources (e.g., on-road deployed sensors, other vehicles, from network, and road users, etc.), evolving the decision making through AI approaches, creating context through interplay of different services distributed across vehicles, edge, and cloud. It is worth highlighting here that various technologies and concepts (of which, some are mentioned above) will enable us attaining the objectives of envisioned Level-5 autonomous driving. However, in this article the focus remains on the adaptive service selection with the objective to ensure the availability of right services when and where required meeting the reliability and delay requirements. The idea is to select the optimal service or a combination of services for different use cases and substitute service(s) that become unavailable with optimally suitable ones. Although the approach is developed for more generic settings of the autonomous driving use cases, for this work, we carried out experiments on a 3.6 km test-road for autonomous driving located at the center of Berlin (refer to Sect. 2 for more details).

The rest of the paper is structured as follows. In Sect. 2, to illustrate the necessity of adaptive sensor selection, we present an autonomous driving testing environment as an enabler of service adaptation. In Sect. 3, we present the adaptive service selection framework with a running example in this testing environment. In Sect. 5, we provide implementation and evaluation details of the prototypical implementation. In Sect. 6, we provide related work related to the adaptive service selection framework. Section 7 concludes the paper with future work.

2 Enriched Information - An Enabler to the Service Adaptation

DigiNet-PS¹ is a research project aiming to create a test-road stretching between Ernst-Reuter-Platz and Brandenburger-Tor in the city center of Berlin. The test-road serves as a playground to deploy and validate prototypes for different use cases in the scope of automated and networked driving. The test environment is built around a 3.65 km long main street, which provides complex traffic situations with end-of-day traffic, government convoys, 2 large roundabouts, 11 traffic signalling systems for vehicles, bicycles, pedestrians, wheelchair users and complex parking space settings. This test area is populated with sensors for parking space detection, environmental conditions (e.g., weather, pollution, and emissions), traffic situation, light status and passers-by information (e.g. pedestrian/cyclist detection).

DigiNet-PS deviates from many classical activities for its philosophy of widening the visibility of vehicles by digitizing (i.e., by deploying the different types of sensors) the roads and consequently creating the perception of road segments

¹ DigiNet-PS project website: <http://dignet-ps.de>.

in the extended roadside units. The sensory data captured through on-road deployed sensors may be fused for studying the correlations between different parameters of the data, processed with AI approaches, and construction of patterns for different decision making instances.

3 Domain Description and Requirements

This section derives the requirements for adaptive service selection in the DigiNet-PS project. For this purpose, we describe an example use case for adaptive sensor selection with the related adaptation requirements.

In our discussions, we will make use of terminology from the domain of opportunistic sensing [2] to abstract from dynamic context recognition approaches. We define the context recognition task as calculating a recognition chain to fulfil a recognition goal. The recognition goal specifies the type of information to be derived via context recognition as well as quality constraints on that information. The recognition chain is calculated by a dynamic context recognition method to fulfil the recognition goal. It consists of sensors providing data as well as processing methods to derive high-level information from the low-level sensor data.

To demonstrate the adaptation requirements, we choose the detection of traffic congestions as an example. The recognition goal is to detect traffic congestions on a specified route. This information could help to reduce traffic jams by suggesting alternative routes or adapting speed before arriving at the congested street.

For detecting traffic jams different types of recognition chains can be calculated. First, camera-based traffic analyzer sensors can provide information about the number of passing vehicles and types of vehicles. Since these sensors are camera based they may suffer from inaccuracies, caused by occlusion or different luminosity values, e.g., under rainy weather conditions. Even though the sensors would cover the same area from different directions of the road, they can provide different sensory data due to the placement or calibration. Due to communication problems, the sensors may become unavailable when this information is requested. Another problem is the number of passing vehicles in a specific location cannot provide information on traffic congestion alone. The data retrieved from sensors should be analyzed and compared with historical data to infer this information. Therefore, additional processing is necessary on top of the retrieved data. Additionally, the system could use information from web services which can provide traffic APIs. For external APIs, there can be a problem of retrieving no up-to-date data in a specific time when an application requires.

For the camera-based sensors, camera position and placement influences the quality of the image recognition (since some parts of the road may be partially covered by trees, or just further away and thus harder to recognize via image processing). Moving vehicles may change the quality provided by the camera even without the camera itself degrading. Additionally, in the running example, even if the accuracy of each camera is known and changes, the distance between the camera and the area to detect traffic instances may change. Traffic instances

that are further away may be harder to detect since with increasing distance the size of the relevant part of the image captured by the camera shrinks and the relative information loss due to image resolution grows. This may lead to lower accuracy of a recognized event. If available, this may be compensated by using higher quality image recognition approaches to detect the situation, which is likely to be computationally more expensive, thus influencing the latency of detection. The quality estimation needs to cover different quality characteristics in combination.

Our aim is to find the right data source(s) to retrieve information about traffic congestion in an optimized way by means of the requested optimization parameters from the application (e.g.; some are in favour of short latency time, for some the accuracy of the retrieved information is more important). Those different ways of getting the same information are called recognition chains in our terminology, which uses sensors and processing of sensory data to create high-level information for the applications. Selection of recognition chains is subject to frequent change due to the inherent dynamics of the testing environment. Therefore, the methods to select related recognition chain should be adaptive to the current situation in order to deal with the changes due to mobility as well as sensing infrastructure.

4 Adaptive Service Selection Framework

In this section, we provide proposed approach to enable adaptive quality-aware service selection in the autonomous driving testing environment. Section 4.1 provides the related concepts definitions to be used hereafter. Section 4.2 presents the overall mechanism from backward chaining to further optimization of selecting the best fitting service.

4.1 Definitions

In this section, we provide brief definitions for recognition goal, recognition chain, service, and quality. The parts of recognition chain consists of different sensors and processing functions, which are modelled as a service. Quality definition is attached to all those services to create aggregated quality of the recognition chain. Formal definition of those terms can be found in our previous work [15].

Recognition Goal (RG): Recognition goal is a definition of the required information with related objects. This includes: identifier of the goal, required input and output as a set of objects, and quality requirements specific to this recognition goal. In the recognition goal, the quality requirements need to be defined to be used later in the optimization function. Quality requirements of recognition goal consist of the identifier for a each single quality property with related maximum and minimum values required to fulfill the goal (if any), the direction to be in favor of (positive when bigger value is better, negative otherwise), and the aggregation method to be used while composing the different parts of a recognition chain (e.g. some quality parameters require to take summation of values in the composition, some require to have the maximum).

Service Model: We define sensor data and processing methods as services in the autonomous driving testing environment. Our service model consists of service description identifier, required sets of inputs and outputs for a service to be executed, and the set of quality properties of a service when it is executed. A single quality parameter consists of the parameter id, type and value of it to use later in a quality calculation of an overall recognition chain.

Recognition Chain (RC): Recognition chain consists of recognition chain identifier, the set of services involved with the identifiers and the execution order in the chain (sequential and parallel execution), the quality properties of the recognition chain that results from aggregation of quality properties of involved services by comparing with the recognition goal. This includes a set of different quality parameters including related id and value after the chaining process.

Quality Model: We define quality in two main categories to fulfill the requirements of adaptive service selection. The first category is quality requirements (QReq), which are attached to the recognition goal. The other category is the quality properties, which are attached to service and recognition chain. The quality of recognition chain is composed by aggregating the quality properties(QProp) of involved services. The calculation of aggregated value for this purpose is provided in Sect. 4.4.

4.2 Quality-Optimal Selection

In this section, an abstract algorithm is provided for the quality-optimal selection of recognition chain to indicate each step in this process. It consists of constructing recognition chains (1), filtering recognition chains whose individual services cannot fulfill the quality requirements of the recognition goal (2), calculating each quality attribute of a recognition chain as a result of quality aggregation (3), filtering recognition chains whose aggregated quality cannot fulfill the quality requirements of the recognition goal (4), selection of best fitting recognition chain after optimization (5). Algorithm 1 returns the best recognition chain after the optimization process. The details of each part in this process are provided in the subsections of this section.

Algorithm 1. Quality-optimal Selection of a Recognition Chain

Require: RG , $serviceDirectory$

```

1:  $QReq \leftarrow RG.getQualityRequirements()$ 
2:  $availableServices \leftarrow loadServices(serviceDirectory)$ 
3:  $RCList \leftarrow construct(RG, availableServices)$ 
4:  $RCListFiltered \leftarrow filter(RCList, QReq)$ 
5: for each RC  $rc_i \in RCListFiltered$  do
6:    $Qprop_i \leftarrow calculateEachAggregatedQualityProperty(rc_i, QReq)$ 
7:    $rc_i \leftarrow rc_i.setQualityProp_i(Qprop_i)$ 
8: end for
9:  $RCListFiltered \leftarrow filter(RCListFiltered, QReq)$ 
10:  $RCListRanked \leftarrow optimize(RCListFiltered, QReq)$ 
11: return  $RCListRanked.getFirst()$ 

```

Constructing recognition chains (1) and filtering the individuals elements in the recognition chain based on recognition goal requirements (2) are presented in Sect. 4.3. Calculating aggregated quality for recognition chain (3) and filtering based on this value against recognition goal (4) are presented in Sect. 4.4. Finally, optimal selection of recognition chain by using quality optimization (5) is presented in Sect. 4.5.

4.3 Recognition Chain Construction and Filtering

In order to create recognition chains from available sensors and processing functions, we map this problem to a service composition problem as we define the elements as a service in the models. One method to solve service composition problems is using a classical AI backward-chaining [14]. Backward chaining starts with the goal state and searches backwards that provide the required inputs until no more unfulfilled input remains.

In order to construct recognition chains from available services, we use backward-chaining and extend the work by Chen et al. [9] to include the quality restrictions in each state of the chaining process. Figure 1 illustrates the general model for composition with quality restrictions. The quality restrictions come from the recognition goal definition (e.g. *Qreq*). While constructing the plans we use the quality requirements retrieved from the request of an application to eliminate the services in the chains which do not satisfy the quality requirements.

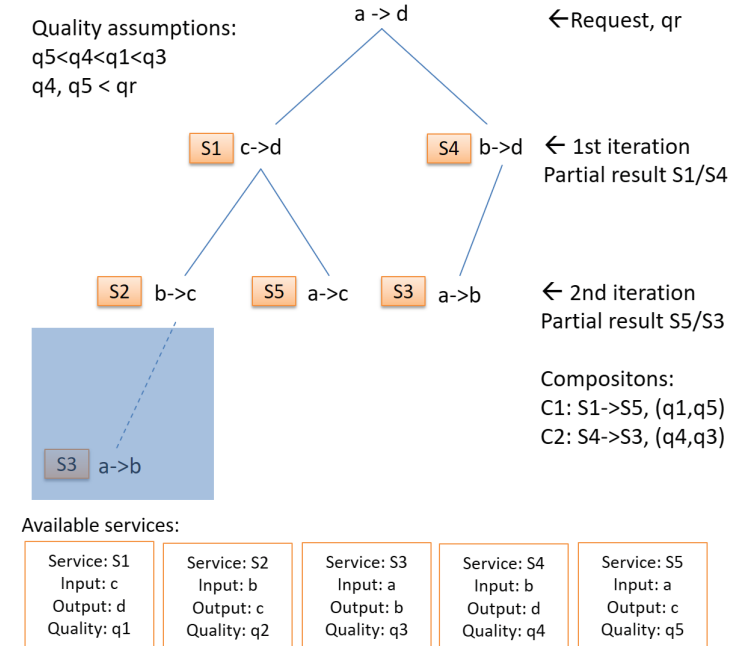


Fig. 1. General backward chaining model with qualities extended from [9]

As can be seen in Fig. 1, there are 5 different services available with the different input and output and quality criteria at the bottom of the figure. In order to reach the goal of (a \rightarrow d) by not considering any quality as stated in [9], the backward search starts to solve the sub-goals by finding the services going to the output of the goal state (S1 and S4). Partial results to the sub-goals are found on the first iteration. Then the search continues to satisfy the output of the services found in the first iteration. Partial results to this sub-goal found on the services S2, S5, and S3. In the second iteration, S5 and S3 already satisfy the given goal state with the resultant compositions C1 and C2 {S1 \rightarrow S5} or {S4 \rightarrow S3} respectively.

We have extended this by including the quality restrictions by assuming different qualities for services as depicted in Fig. 1 upper left-corner. Based on this assumption q4 and q5 cannot satisfy the requested quality for the goal which makes the found composites obsolete. Therefore, the back-chaining needs to continue until to find a solution of the sub-goal which also satisfy the quality requirement. Therefore, the last iterations state is saved to continue in order to find a satisfying service for the sub-goals. In the third iteration, S3 is found to satisfy the sub-goal and as its input is a, therefore the overall goal is also satisfied with the constructed composition S1 \rightarrow S2 \rightarrow S3. In the chain construction, we eliminated the services which cannot fulfil the recognition goal requirement. However, when there is no violation of requirements we keep the constructed chains to further optimize based on overall chain quality. The overall chain quality calculation is provided in Sect. 4.4.

Within the guidance of extended backward chaining, to create a recognition chain, we search for available data sources and processing method services in the distributed service repository. When any service violates the quality requirement from the recognition goal, it is eliminated and the chain is stopped to go further iteration from this branch. If there is no violation, we keep the constructed chains with the iteration order to further calculate overall chain quality presented in Sect. 4.4.

4.4 Quality Aggregation and Filtering:

In order to select the best recognition chain amongst the ones created as a result of backward chaining explained in Sect. 4.3, we need to first calculate the quality of the overall chain (*Qeval* in recognition chain description). For this purpose, each quality parameter in the recognition goal (*Qreq*) should be iterated and based on the type of the quality parameter, related aggregation method, a direction of optimization and minimum and maximum boundaries of the parameter should be applied. For example, for latency in a chain of services can be aggregated by using a SUM operator to sum of all consisting services in the chain whereas for the accuracy of information MIN operator should be applied for the aggregation of different accuracy value of information to serve as a boundary. A direction of a quality parameter indicates to the favour of smaller or larger values. In our quality description, negative direction indicates smaller values are favourable. Maximum and minimum boundaries from the

recognition goal description in this phase should be also taken into account for the calculation. Although single services do not violate the boundaries as a result of extended backward chaining process presented in the previous section, the aggregated quality value of a recognition chain should remain within the boundaries.

Algorithm 2 shows a quality calculation algorithm for a recognition chain based on each quality attribute as a result of aggregation. Algorithm 2 returns the aggregated quality values for each quality parameter in the recognition chain. As a result, a list of quality properties for a recognition chain including related id and quality value is found and assigned to the recognition chain quality property (*rcQProp*).

Algorithm 2. Quality Calculation for a Recognition Chain

Require: *RC*, *RG*

```

1: QReq  $\leftarrow$  RG.getQualityRequirements()
2: SSer  $\leftarrow$  RC.getServices();
3: Qprop  $\leftarrow$  new Listi
4: for each quality requirement qri  $\in$  QReq do
5:   qrID  $\leftarrow$  qri.getID()
6:   qrType  $\leftarrow$  qri.getAggregationType()
7:   qrMax  $\leftarrow$  qri.getMinimum()
8:   qrMin  $\leftarrow$  qri.getMaximum()
9:   for each service sj  $\in$  SSer do
10:    sQProps  $\leftarrow$  sj.getQualityProperties()
11:    qProp  $\leftarrow$  0.0
12:    for each prop pk  $\in$  sQProps do
13:      pID  $\leftarrow$  pk.getID()
14:      if pID equals qrID then
15:        pValue  $\leftarrow$  pk.getValue()
16:        switch (qrType)
17:          case: "SUM"
18:            qProp += pValue break
19:          case: "MAX"
20:            qProp = MAX{qProp,pValue} break
21:          case: "MIN"
22:            qProp = MIN{qProp,pValue} break
23:          case: "MULT"
24:            qProp *= pValue break
25:        end if
26:      end for
27:    end for
28:    if (qProp  $\neq$  0.0)  $\wedge$  (qProp  $\leq$  qrMax)  $\wedge$  (qProp  $\geq$  qrMin) then
29:      rcQProp  $\leftarrow$  qProp, pID
30:      SetQprop.add(rcQProp);
31:    end if
32:  end for
33: return SetQprop

```

After each quality value is calculated separately for a recognition chain, they are checked against the boundaries of the $Qreq$. If there is a violation of any aggregated quality property of a recognition chain based on the recognition goal, this recognition chain is discarded. If there is no violation of any quality parameter in the aggregated values, the resultant quality values of recognition chains are kept for further optimized to use different quality properties in combination to select the best recognition chain. This process is explained in Sect. 4.5.

4.5 Quality-Optimal Selection

In order to use different quality values of recognition chains in a single optimization function, the values should be used independent of their metric. This requires to normalize the values in an interval preferably between 0 and 1. The main problem is having unbounded values for some of the quality properties of the services. If there is no requirement from recognition goal for this property, the boundaries of this value should be found based on the observed values. For this purpose, we have implemented different statistical techniques (tanh, sigmoid, z-score, logistic regression, pnorm) to find out which one is suitable for the observed values for each different quality property to find a normally distributed scale between 0 and 1. The behaviour of how different techniques work on the observed values for normalization is depicted in Fig. 2.

As can be seen from Fig. 2, except tanh and softmax, the other methods provide close to equal distribution for the normalization. For each observed quality variable, we have used different technique fitting the with the related sample mean and standard deviation calculated based on the existing observed values for each quality property. After the normalization phase, we are able to use them as a weighted sum of different quality parameters in an optimization function to calculate the overall quality of a recognition chain.

The algorithm decides for an optimum recognition chain to select based on the recognition goal. The mathematical formulation is given as a Linear Programming optimization problem. The constraint for the objective function can be formulated as maximization function as defined in Eq. 1.

$$\underset{rc}{\text{maximize}} \quad \sum_{i=1}^n w_i x_i(rc), \quad \text{where} \quad \sum_{i=1}^n w_i(rc) = 1, \quad 0 \leq x_i(rc) \leq 1 \quad (1)$$

This function assumes that there are quality characteristics x_0 to x_n are normalized values which can be derived from a recognition chain rc after the normalization and weights for these quality characteristics w_0 to w_n . The goal is to find a recognition chain that maximizes the weighted sum of these quality characteristics. We have applied Linear Programming (LP) to solve the equation in order to find the optimal values for each quality parameter in the optimization function.

After finding the optimal values, we have used those values to calculate the distance for each recognition chain from the optimal values. We sum up the

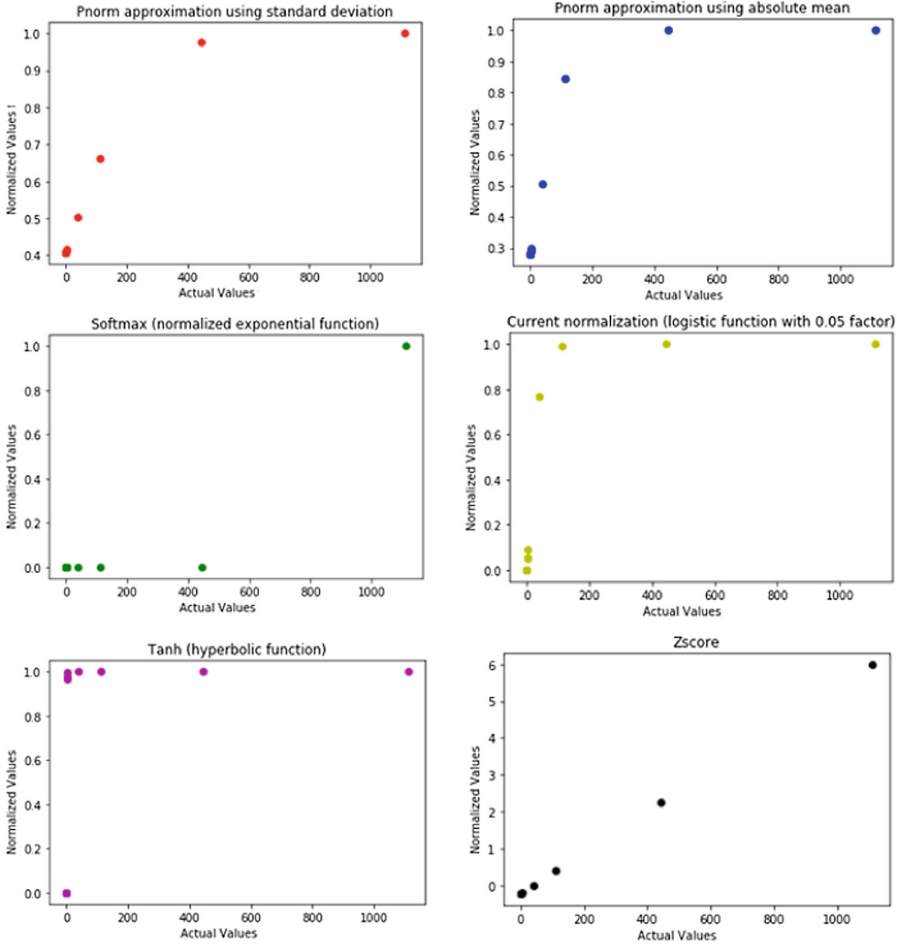


Fig. 2. Different statistical methods for normalization

values for each separate distance and ranked the recognition chains based on the minimum distance from the optimal value.

$$\underset{rc_{dist}}{\text{minimize}} \sum_{i=1}^n w_i(x_{opt} - x_i(rc)). \tag{2}$$

where x_{opt} is the optimal value for each quality property found as a result of solving LP in the Eq. 1. As a result of Eq. 2 to calculate the distance, we have a ranked list of optimal recognition chains.

5 Implementation and Evaluation

In this section, we discuss how we implemented the presented approach in the previous section in terms of specific technologies and evaluated the calculation of selecting the optimal recognition chain and adaptation of it during run time by using a running example.

As proposed in [10], the architecture for adaptive service selection is agent-based and implemented using JIAC V, a Java-based Intelligent Agent Framework² to handle run time changes as well as providing communication and discovery capabilities of services deployed in a distributed environment [8]. To handle the runtime changes, different agents are implemented first for monitoring the sensors and processing functions, second for keeping track of deployed recognition chains as a result of service composition and quality optimization. We have implemented another agent to store recognition goal and quality requirements for finding suitable recognition chain.

The statistical techniques for the optimization are implemented in Python to compare the results of each method for normal distribution. Additionally for solving the LP optimization problem, we have used CPLEX³ implemented in Python.

For evaluation purposes, service agents have been implemented that mimic the behavior of real services with related quality properties. In the next section, we present a running example to illustrate the details for evaluation.

5.1 Running Example

In this section, we present a running example to explain the implementation details more intuitively. Our running example is taken from the autonomous driving domain. The adaptive system is part of the control system of an autonomous vehicle. In addition, we assume that the vehicle can access external sensors (e.g., deployed on the road-side) in addition to its on-board sensors, to improve context detection. Thus, it has a need for sensor selection as the relevant sensors change with the position of the vehicle and the planned route.

In the running example, we focus on the detection of traffic congestions on a planned route. Our approach is responsible for finding and maintaining recognition chains that can provide this information. For this purpose, request description as a recognition goal contains a route r along which traffic congestions should be detected. For the sake of simplicity we represent r as a set of streets. The recognition goal specifies that the target of the recognition process is a heat-map of traffic congestion risk, represented as a mapping from the streets in r to real numbers between 0 and 1. This recognition goal is accompanied by an objective function that is a weighted sum of two quality requirements the latency of the resulting recognition chain (i.e., the time it takes to detect a change in traffic

² JIAC V, <http://www.jiac.de/agent-frameworks/jiac-v/> [Accessed: May 30, 2019].

³ CPLEX Optimizer: <https://www.ibm.com/analytics/cplex-optimizer> [Accessed: May 30, 2019].

jam risk) and the accuracy of the detected heat-map. The objective function is maximized, meaning that accuracy is maximized and latency is minimized. This may require a trade-off if the more accurate detection also takes longer.

In order to construct recognition chains from available sources, describes the sensors available to calculate traffic congestions from. These include roadside information sources, like vehicle counters and cameras, and online information sources from external online map APIs. The processing functions include methods to derive traffic congestion information from low-level sensors like vehicle counters (*VehicleCount2Congestion*) and cameras (*Camera2Congestion*). It also contains a function *Single2Heatmap* that merges a set of processing functions for single streets into the heat-map for all streets on the route specified in the recognition goal. The quality attributes are given by two prediction functions that are able to predict latency and accuracy for any sensor and processing function. This information is partially derived from the *RCPPerformance* information, which has been collected from running recognition chains.

Backward chaining with quality extension combines the sensors and processing functions and updates the *CurrentRecognitionChain*. The expected outcome of this phase in the running example is a recognition chain that uses service *Single2Heatmap* as the last service and has dedicated processing functions for each street in r . These recognition chains either use the *TrafficAPI* or combinations of *VehicleCount* and *VehicleCount2Congestion* or *Camera* and *Camera2Congestion*. At the end of this phase, there are recognition chains available to further optimize and finding the best one based on the recognition goal.

In the last phase as optimization, our approach looks related quality parameters and find the best matching chain based on the recognition goal. This process is done by solving LP optimization problem and ranking the fitting recognition chains based on the distance from an optimal solution. Based on this, the recognition chain is selected and updated in case of any change on the measured quality parameters.

In order to create necessary chain for this purpose and update in case of any changes in the services, we have provided the setup and evaluation explained in the next section.

5.2 Setup and Evaluation

To evaluate the effectiveness of proposed approach we designed two experiments targeting the overall performance enabling adaptive service selection at runtime and the correctness of the results after the optimization process. The first experiment is measuring the performance of the algorithm, and the second experiment is testing its adaptive behaviour in case of changes.

The experiment setup consists of a virtual machine (VM) with a two core Intel Xeon Processor E7 (2.27GHz) and 10 GB of RAM, running an UBUNTU 16.04 LTS operating system. The VM has a Apache Kafka⁴ instance running

⁴ Apache Kafka: <https://kafka.apache.org/> [Accessed: May 30, 2019].

on it to retrieve sensor data used in the services. As Apache Kafka is based on publish-subscribe mechanism, the changes about the service quality is forwarded directly to the agent subscribe to related topic for this specific service.

Experiment 1: To test the performance of the algorithm for the optimization on top of backward chaining, we have tested services with 10, 20, 40, 50, 100 service deployment setup. For the performance, we have setup warm-up iterations with a number of 100 and then run the iterations with a number of 100 to calculate the average time. The results are depicted in Fig. 3.

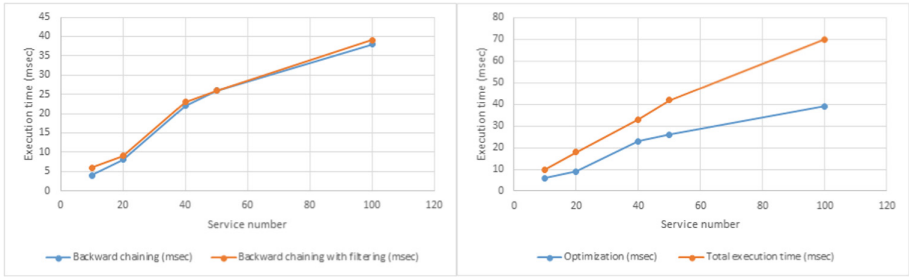


Fig. 3. Execution times with different service numbers

As can be seen in Fig. 3 (left), the execution time of backward chaining and backward chaining with quality filtering is close to each other with slightly in favor of backward chaining in the low number of services and the opposite for the high number of services. Figure 3 (right) shows the required optimization time with the total for finding the optimal recognition chain. Although the optimization increases the total execution time, this can be compensated if it can decrease the failure rate of a service.

Experiment 2: To test the adaptive behaviour of the system, we have created the setup with 10, 20, 40, 50, 100 service deployment setup by randomly making some services unavailable. We have measured the success rate of the mechanism in case of changes with the non-adaptive version. For this purpose, we have compared the output of the required recognition goal with the result of recognition chains. The results are depicted in Fig. 4 for time required for adaptive behaviour of the system with the success rate.

As can be seen in Fig. 4 (left), success rate for adaptive selection is higher than the non-adaptive version with a execution time difference (right) trade-off. Although the adaptive behaviour increase the execution time, based on the results to increase success rate, it is still in an acceptable range for some conditions in the test road.

The experiments show promising results for the increase on success rate, however the setup should be extended by deploying more services to different

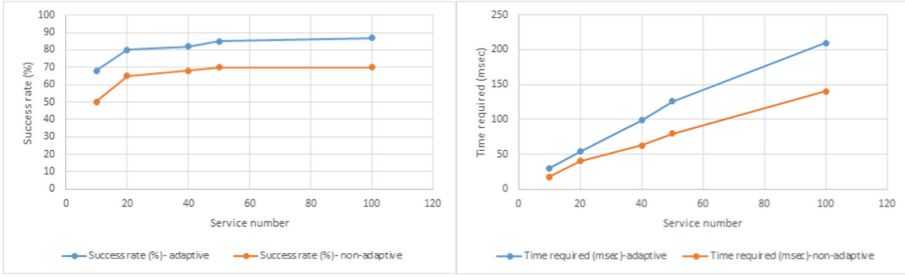


Fig. 4. Success rate with the execution times for different services

nodes to make more realistic for different use cases in autonomous driving testing environment.

6 Related Work

Quality estimation with multiple conflicting criteria is maturely researched in the scope of multi-objective optimization especially in web services domain [11]. The area is still challenging as the context recognition is not only influenced by the services but also properties of the digital parts of the devices and sensors. The influence of methods to process device/sensor data is also important to be taken into consideration as they can affect the overall quality of the retrieved information. Our main contribution is to take into account different parameters which affect overall quality while constructing service chains and adapting the chain to the changing conditions. We enabled this by extending backward-chaining approach by Chen et al. [9].

Opportunity Framework [2] is as a reference implementation of opportunistic sensing for human activity and context recognition. In Opportunity, a numeric value and its degree of fulfilment are used as a quality metric for the accuracy of recognized activity to dynamically configure a recognition chain [2]. This metric does not take into account other quality metrics like the frequency of detection or the costs of detection. Villalonga et al. presented guidelines to match quality parameters used to assess the quality of context in a generic way and mathematical formulations for the conversion in activity recognition domain [12]. However, those conversion functions do not take into account the requirements from the application side. Manzoor et al. proposed a model for processing different QoC metrics by differentiating QoC as subjective and objective view depending on the application requirements [13]. Their approach is using one entity (context object) to calculate those quality parameters and does not take into account the combination of different context objects (e.g. information providers) using those QoC parameters. Different than this work, we have addressed the combination of different quality parameters (QoS, QoD) from different data sources affecting overall quality (QoC).

Everything-as-a-Service (XaaS) [6] concept has been in cloud computing for many years to use available hardware and software resources in an efficient way. Major service models proposed under XaaS, like Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS), were limited with cloud computing at the early stage. With the evolvement of IoT, new service models are proposed. One of those models is Sensing-as-a-Service (S2aaS) [3, 4] supporting the idea of data exchange between the data owners and data consumers in IoT. Similarly, the Context-as-a-Service (CoaaS) model [5] has evolved as an abstraction layer on top of the data that can be retrieved from different resources via IoT and aims at addressing the gap between context consumers and producers to provide interoperable data exchange [7]. In DigiNet-PS, we have inspired by XaaS to bring a solution addressing this gap by modelling and developing every part of the information sources as well as processing methods of those sources as a service. This gives us an opportunity to use mature techniques from the web services domain with many different optimization approaches.

7 Discussion and Conclusion

In this paper, we have provided adaptive service selection framework consisting of selection mechanism and optimization based on different quality criteria of the services. We have evaluated the result of the mechanism from correctness of the output of the service and performance of overall execution in an autonomous driving testing environment data set inspired by the real conditions in the scope of traffic congestion use case. The results show that although the adaptive behaviour increase the execution time, it also relatively increase the success rate in case of any changes in the service quality. The execution time of the mechanism is still in an acceptable range for some conditions/use cases in the test road.

For future work, we aim to test and evaluate the initial result of the framework with different use cases in different physical deployment by using edge computing. Especially, using the edge to minimize communication delay and thus improves the overall quality of the recognition chains could improve the latency more for time-critical actions required in the testing environment.


References

1. 3GPP specification for 5G. <https://www.3gpp.org/release-15>. Accessed 30 May 2019
2. Kurz, M., Hoelzl, G., Ferscha, A., et al.: The OPPORTUNITY framework and data processing ecosystem for opportunistic activity and context recognition. *Int. J. Sens. Wirel. Commun. Control* **1**(2), 102–125 (2011)
3. Perera, C., Zaslavsky, A., Christen, P., Georgakopoulos, D.: Sensing as a service model for smart cities supported by Internet of Things. In: *Proceedings of the Transactions ETT*, p. 113 (2013)

4. Zaslavsky, A., Perera, C., Georgakopoulos, D.: Sensing as a service and big data. In: International Conference on Advances in Cloud Computing (ACC 2012), Bangalore, India, pp. 21–29 (2012)
5. Wagner, M., Reichle, R., Geihs, K.: Context as a service - requirements, design and middleware support. In: 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), pp. 220–225 (2011)
6. Banerjee, P., et al.: Everything as a service: powering the new information economy. *Computer* **44**(3), 36–43 (2011)
7. Moore, P., Xhafa, F., Barolli, L.: Context-as-a-service: a service model for cloud-based systems. In: 2014 Eighth International Conference on Complex, Intelligent and Software Intensive Systems, pp. 379–385. IEEE (2014)
8. Lützenberger, M., et al.: A multi-agent approach to professional software engineering. In: Cossentino, M., El Fallah Seghrouchni, A., Winikoff, M. (eds.) EMAS 2013. LNCS (LNAI), vol. 8245, pp. 156–175. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-45343-4_9
9. Chen, N., Cardozo, N., Clarke, S.: Goal-driven service composition in mobile and pervasive computing. *IEEE Trans. Serv. Comput.* **11**(1), 49–62 (2018)
10. Eryilmaz, E., Trollmann, F., Albayrak, S.: An architecture for dynamic context recognition in an autonomous driving testing environment. In: 2018 IEEE 11th Conference on Service-Oriented Computing and Applications (SOCA), pp. 9–16, November 2018
11. Strunk, A.: QoS-aware service composition: a survey. In: Proceeding of 8th European Conference on Web Services (ECOWS), pp. 67–74 (2010)
12. Villalonga, C., Roggen, D., Lombriser, C., Zappi, P., Tröster, G.: Bringing quality of context into wearable human activity recognition systems. In: Rothermel, K., Fritsch, D., Blochinger, W., Dürr, F. (eds.) QuaCon 2009. LNCS, vol. 5786, pp. 164–173. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04559-2_15
13. Manzoor, A., Truong, H.-L., Dustdar, S.: Quality of context: models and applications for contextaware systems in pervasive environments. *Knowl. Eng. Rev.* **29**(2), 154–170 (2014)
14. Hibner, A., Zielinski, K.: Semantic-based dynamic service composition and adaptation. In: Proceedings of the IEEE Services Congress, pp. 213–220 (2007)
15. Eryilmaz, E., Trollmann, F., Albayrak, S.: Quality-aware service selection approach for adaptive context recognition in IoT. In: Proceedings of the 9th International Conference on the Internet of Things (IOT 2019). ACM, Bilboa (2019, to appear)



Discovering User Location Semantics Using Mobile Notification Handling Behaviour

Andreas Komninos^{1,2}(✉) , Ioulia Simou², Elton Frengkou²,
and John Garofalakis^{1,2}

¹ University of Patras, 26504 Rio, Greece

² Computer Technology Institute and Press “Diophantos”, 26504 Rio, Greece
{akomninos,simo,frengkou,garofala}@ceid.upatras.gr

Abstract. We analyse data from a longitudinal study of 44 participants, including notification handling, device state and location information. We demonstrate that it is possible to semantically label a user’s location based on their notification handling behaviour, even when location coordinates are obfuscated so as not to precisely match known venue locations. Privacy-preserving semantic labelling of a user’s location can be useful for the contextually-relevant handling of interruptions and service delivery on mobile devices.

Keywords: Interruption management · Mobile notifications · Semantic location labelling

1 Introduction

As users of mobile devices roam through urban environments, a wealth of data can be collected from their devices about their current whereabouts and activities. While it is relatively easy to obtain the location of a user, within a given accuracy estimate (e.g. through GPS, connection to Wi-Fi or 4G networks), a harder task is to assign semantics to the user’s location. The typical method of resolving this, is by comparing the user’s coordinates against a database of known locations, and there are several commercial services that offer this type of information (e.g. Google Places API). Therefore, given a user’s location coordinates, it is relatively easy to obtain the venue that a user might currently be at, and therefore to infer their current activity (e.g., they are at Cinema X, and thus quite likely watching a movie). This knowledge is valuable for the purpose of offering contextually relevant notification handling to users. Currently, users are left on their own in terms of how they might manage notifications under different contexts [3]. However, automatic notification management can offer opportunities for a better and more socially aware mobile use experience [2]. Taking the cinema example, a device could automatically suppress incoming notifications which are not relevant at the current location (e.g. [11]) or automatically set the device ringer mode to silent for the duration of the user’s stay at that location.

There are several confounding factors to being able to achieve this goal. First, user location coordinates might not be available, or accurate enough to provide a reasonable estimate of venue (e.g. the user might be indoors, or the user might be connected to a sparse 4G network only). Further, the user might be mobile and therefore rapidly moving across venues, hence a continuous lookup of the user’s location is required, expending device power and network bandwidth. Even more, for services such as this to work, the user’s location needs to be sent to a remote server, potentially compromising user privacy.

As discussed in existing literature, users receive a significant volume of notifications during the day, from on-device events (e.g. network availability, battery status) and external services (e.g. instant messaging), which can reach several hundreds [12]. These events can become opportune moments for assessing the user’s location. The user behaviour in handling these notification events can vary significantly across time (e.g. [8]), and we can assume that the behavioural choices are influenced by the location context and semantics as well, even though there is no previous literature to investigate this. For example, while watching a movie at the cinema, the user might take longer to notice an incoming notification since their device will probably be set to “silent mode” and tucked away, or even if they do, they might chose to ignore it until the show is over.

In this paper, we explore the use of notification handling behaviour and device state information, as an additional source of information for overcoming problems with user coordinate availability and accuracy. Using supervised machine learning algorithms on a dataset of notification and location samples from several users, we predict user location semantics and demonstrate that notification handling behaviour can overcome the problem of location accuracy.

2 Related Work

Discovering location semantics is the research effort directed towards assigning categorical labels (e.g. “Home”, “School”, “Shop”) to venues represented in a dataset with at least a set of coordinates (latitude, longitude) and optionally a given name (e.g. “Mike’s cafe”). Location semantics are important for a range of location based services, such as point-of-interest (POI) search and recommendation. Commercial applications such as Google Maps, Foursquare and Tripadvisor maintain large databases of POIs, relying largely on users adding and/or modifying these. One issue with this approach is that represented venues are not always correctly semantically labelled by the users, and also the reliance on user effort means that many real-world POIs may be often left out of the service. Previous research has frequently focused on the automatic semantic labelling of locations, with a variety of means. For example, in [6], check-ins from social networks (Twitter) were used to identify users’ home locations, with good accuracy. In [9], data from location diary studies was used to build a model to automatically infer user home and office locations, using GPS traces as an input, resulting in reasonable performance for both categories (100% and 66%). In [5], check-in data (e.g. number of visitors, diurnal distribution, stay time etc.) was again used

to predict venue labels across 8 categories, however with mixed results across the different categories (F-score between 54%–92%). In [7], a spatiotemporal topic model was used to leverage location “tags” left by users, in order to determine the location category, with an average accuracy $\approx 60\%$.

Other studies have leveraged sensor data in addition to other contextual information for semantic place labelling. In [15], GPS, accelerometer, Bluetooth and Wi-Fi data were used amongst others to achieve an accuracy $\approx 75\%$ across 10 different location categories. Similar results are obtained in [13], where labelled and unlabelled data are used to implement a semi-supervised learning approach to predict across 9 categories. In [4], 11 categories are predicted from, using features related to stay time, device battery, applications used, user current activity etc. Results show an accuracy of $\approx 55\%$ with a range of classifiers.

There are some common themes in the previous literature, which can be identified. First, where multiple classifiers have been used (e.g. decision trees, SVMs, random forests), the results do not seem to vary significantly. Most often, it is the type and number of features introduced to the model which have the most impact. Secondly, a larger number of categories makes the likelihood of misclassifications higher. Both in [4] and in other work such as [10], it is demonstrated that a less fine-grained categorisation approach improves results significantly (e.g from $\approx 65\%$ with 10 categories, to $\approx 89\%$ when these are collapsed into 3). Another issue is that, as demonstrated in most papers (e.g. [4]), there is a significant class imbalance in the datasets used. This is somewhat problematic since in all reviewed works apart from [5], the measure of accuracy is used, which is heavily influenced by the prevalence of certain categories [1]. Hence, comparisons with the performance of these previous approaches is done with some hesitation.

To the best of our knowledge, the use of notification handling behaviour as a feature for semantic place labelling has not been investigated in the past. Hence the goal of our paper is to explore how this information can be used for the task of semantic place labelling.

3 Study Methodology

3.1 Apparatus and Participants

We developed a UI-less notification logging application for Android devices, which runs unobtrusively on the device as a background service. Using the Android NotificationListener service, which allows the capture of issued notification details, as well as various other Android APIs (e.g. PowerManager, DisplayManager), we collected features about the notifications and the user’s device state at the time of issue. We also exploited the Google Places API to retrieve details about the user’s presumed location at the time of notification issue. This API requests the user’s location coordinates, and returns a list of likely places where the user is located, along with a confidence level. We logged the place which had the highest confidence value. The data features collected are discussed in detail in Sect. 3.3. All data was uploaded to a remote server at frequent intervals during the day, provided the user had wi-fi connectivity.

A call for participation was issued to undergraduate students at our local university. The application was installed on their device, a consent form was signed and participants were instructed that they could quit the study at any time. The study automatically ended after 3 months of use. They were requested to leave location services enabled on their device for the duration of the study, although we did not enforce this condition. In total, 44 participants took part in the study (26 female). From this set of participants, we excluded several participants who participated for fewer than 10 days and who provided fewer than 50 notification log entries, resulting in a subset of 31 participants. Participants provided data that spanned an average of 30.87 days ($sd = 16.15$, $min = 13$, $max = 84$).

3.2 Dataset Preparation

In total we collected 204,074 notifications from the users. In the dataset, we noticed that a significant number of notifications (38,400) were issued by the system and immediately dismissed. This phenomenon was observed for all users, although for some users the proportion of such notifications was unusually large. We are not certain why this happens. Further investigation of the package name showed that some system applications might be issuing such notifications (perhaps as a means of interprocess communication), although it might be the case that a user is also manually quickly dismissing some notifications (within the resolution of 1 sec, used to capture timestamps for logging). We decided to exclude such notifications from the dataset. Further, we removed from the dataset all notifications for which the “flag” feature values indicated that they were ongoing events and not user-dismissable (e.g. an ongoing phonecall or download). These notifications are automatically dismissed by the system and hence offer no value to our research goal. From the remaining notifications, a significant number did not contain location information, since the user’s location services might have been switched off at the time, or the service might not have been available. We also excluded these from the dataset. After these exclusions, the dataset contained 59,221 user-dismissed notifications with location details.

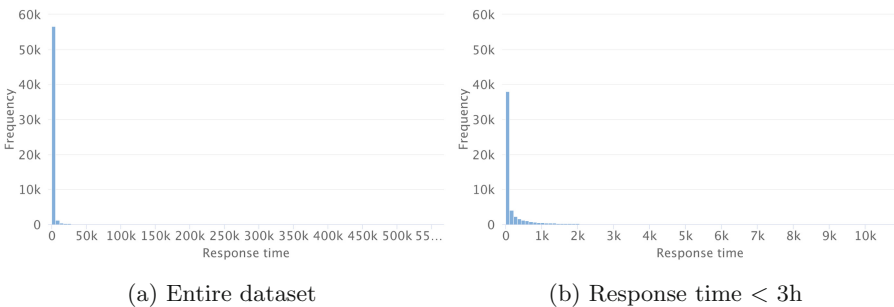


Fig. 1. Distribution of response time to notifications (100 bins)

Examining the pruned dataset, we observed that the average response time to notifications is 1,366.93 s ($sd = 11,255.82$), with a maximum response time of 562,302 s. A histogram of response time to notifications shows a power-law distribution (Fig. 1). Based on this observation, we limited the dataset to only notifications that were attended to within 3 h of issue, resulting in 57,737 notifications (97.5% of the original dataset). As can be seen, even after culling the dataset further, the distribution of response times to notifications maintains a power-law shape. This finding is consistent with previous works such as [8].

3.3 Dataset Features

To address the problem at hand, we used raw and synthetic features obtained from the user’s device. To begin, the raw data features collected from users are shown in Table 1.

From these raw features we synthesized a further set of features, to create the final dataset to be used for prediction, as shown in Table 2. Notably, we used the current device ringer mode and programmed notification modalities (custom or default) to determine the true modalities used to deliver the notification, as per [8]. Further, a place can belong to multiple categories. These are reported in a non-ordered list by Google, ostensibly therefore the order of appearance shows the prevalence of a category type (e.g. “Bar, Restaurant, Cafe” shows that a place is primarily of type “Bar”, but also functions as a restaurant and cafe). We therefore extract the primary category of a venue. In doing so, we observed that many places included the vague category “Point of Interest”. Hence, where this was the primary category, it was replaced by the immediately subsequent category type.

Another note here relates to Google’s list of categories, where 127 different categories are listed. Predicting on 127 category classes is possible, but presents an unnecessary complexity to the problem, as many venue categories are quite similar in nature and it can be expected that a user will exhibit similar behavioural patterns in these. For example, “Church” and “Mosque” are both places of worship, where devices are typically kept on silent, and users do not readily engage in notification handling. We therefore attempted to group the individual categories into larger sets, as per Table 3. Ultimately, we assigned to each place the super-category to which it belongs, based on its primary category type. An exception to this were the “Miscellaneous” and “Entertainment areas” categories, since for these the user behaviour might be quite different depending on conditions (e.g. a user probably can’t notice a notification in a night club as easily as in a cafe), hence for these we used the primary categories ungrouped. As a result, we find that the user notifications were issued at 24 distinct place categories and distributed unevenly (Table 3, non-grouped primary categories capitalised). Finally, it’s important to note that the location coordinates collected by our app, are not the user’s actual coordinates, but the coordinates of the venue that is the user’s most likely current place, as reported back by Google’s API. We do not store the user’s actual location coordinates for privacy reasons.

Table 1. Raw data features collected

<i>Notification details</i>	
Time posted	Unix timestamp of notification issue
Time dismissed	Unix timestamp of notification dismissal
Package name	Application that created the notification
Sound	Whether the notification was programmed to issue a custom sound alert
LED	Whether the notification was programmed to issue a custom LED blink pattern
Vibration	Whether the notification was programmed to issue a custom vibration pattern
DefaultSound	Whether the notification was programmed to use the default sound alert
DefaultLED	Whether the notification was programmed to use the default LED blink pattern
DefaultVibration	Whether the notification was programmed to use the default vibration pattern
Priority	The notification priority category
Notification flags	Additional information about the notification
<i>Device state</i>	
Ringer mode	The current device ringer mode (silent, vibrate only, full)
Idle state	Whether the device is in an idle state
Interactive state	Whether the device is in a state ready to interact with the user (screen on, processor awake)
Lockscreen notifications allowed	Whether notifications are visible from the user's lock screen
<i>Location details</i>	
Place name	Name of the most likely current place
Place categories	The categories assigned to the most likely current place
Confidence	Confidence of reporting the most likely current place
Latitude	Decimal coordinates of the most likely current place
Longitude	Decimal coordinates of the most likely current place

Table 2. Final feature set

<i>Notification details</i>		
Response time	Time dismissed - time posted	Synthetic
Hour issued	Hour of day at notification issue [0–23]	Synthetic
Day of week issued	Day of week at notification issue [1–7]	Synthetic
Had Sound	Whether the notification was issued with a sound	Synthetic
Had LED	Whether the notification was issued with a LED blinking pattern	Synthetic
Had Vibration	Whether the notification was issued with a vibration pattern	Synthetic
Priority	The notification priority category	Raw
<i>Device state</i>		
Idle state	Whether the device is in an idle state	Raw
Interactive state	Whether the device is in a state ready to interact with the user (screen on, processor awake)	Raw
Lockscreen notifications allowed	Whether notifications are visible from the user’s lock screen	Raw
<i>Location details</i>		
Place category	The primary place category	Synthetic
Latitude	Decimal coordinates of the most likely current place	Raw
Longitude	Decimal coordinates of the most likely current place	Raw

As can be seen in Fig. 2, users receive a varying amount of notifications throughout the day. The distribution is similar to that reported in previous literature, such as [4]. More importantly, we note that the diurnal distribution varies pronouncedly for only a few categories, whereas for other categories, it remains rather consistent. This is an expected result, since different venue types exhibit different diurnal visitation patterns [5]. Further, we note the distribution of response times to various notifications on a hourly basis (Fig. 3a). The pattern is similar to the findings in [8], showing the distinct user behaviour in handling notifications throughout the day. Distinct response time averages are also noted across the categories (Fig. 3b), demonstrating that attentiveness to the device is likely related to device ringer mode and current user activity (Fig. 4).

Table 3. Grouped place categories

Category group	Categories	Samples
Accommodation	Campground, Lodging, Room, Rv Park	1,350
Address	Administrative Area Level 1, Administrative Area Level 2, Administrative Area Level 3, Country, Geocode, Locality, Political, Post Box, Postal Code, Postal Code Prefix, Postal Town, Street Address, Sublocality, Sublocality Level 1, Sublocality Level 2, Sublocality Level 3, Sublocality Level 4, Sublocality Level 5, Synthetic Geocode	86
Civil services	City Hall, Courthouse, Embassy, Fire Station, Local Government Office, Police, Post Office	89
Contractors	Electrician, General Contractor, Moving Company, Painter, Plumber, Roofing Contractor	76
Education	Library, School, University	11,996
Entertainment areas	Amusement Park, Aquarium, Bar, Bowling Alley, Cafe, Casino, Gym, Movie Theater, Museum, Night Club, Restaurant, Stadium, Zoo	11,157
Financial services	Bank, Atm, Finance	93
Healthcare	Dentist, Doctor, Health, Hospital, Physiotherapist	617
Miscellaneous	Establishment, Floor, Other, Point Of Interest, Premise, Subpremise	18,347
Outdoor areas	Colloquial Area, Natural Feature, Neighborhood, Park, Parking, Route	516
Personal care	Beauty Salon, Hair Care, Spa	1,104
Place Of worship	Cemetery, Church, Hindu Temple, Mosque, Place Of Worship, Synagogue	758
Professional services	Lawyer, Accounting, Car Dealer, Car Rental, Car Repair, Car Wash, Funeral Home, Insurance Agency, Laundry, Locksmith, Real Estate Agency, Storage, Travel Agency, Veterinary Care	659
Public transport	Airport, Bus Station, Intersection, Subway Station, Taxi Stand, Train Station, Transit Station	580
Shopping	Art Gallery, Bakery, Bicycle Store, Book Store, Clothing Store, Convenience Store, Department Store, Electronics Store, Florist, Food, Furniture Store, Gas Station, Grocery Or Supermarket, Hardware Store, Home Goods Store, Jewelry Store, Liquor Store, Meal Delivery, Meal Takeaway, Movie Rental, Pet Store, Pharmacy, Shoe Store, Shopping Mall, Store	10,309

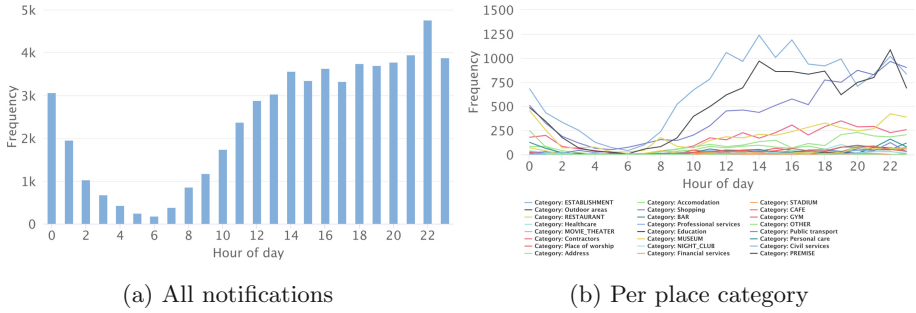


Fig. 2. Diurnal distribution of notifications

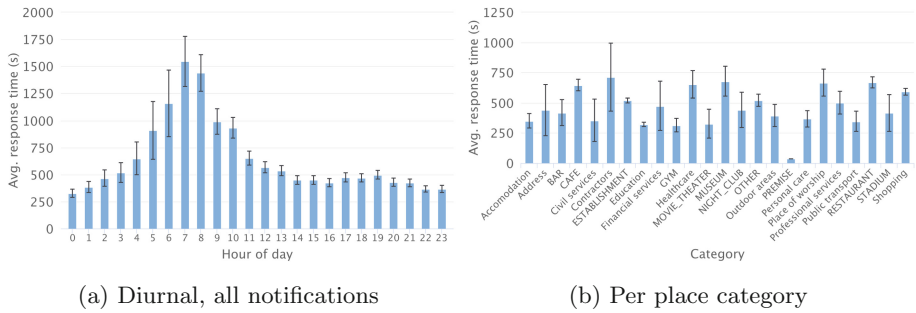


Fig. 3. Distribution of response time to notifications

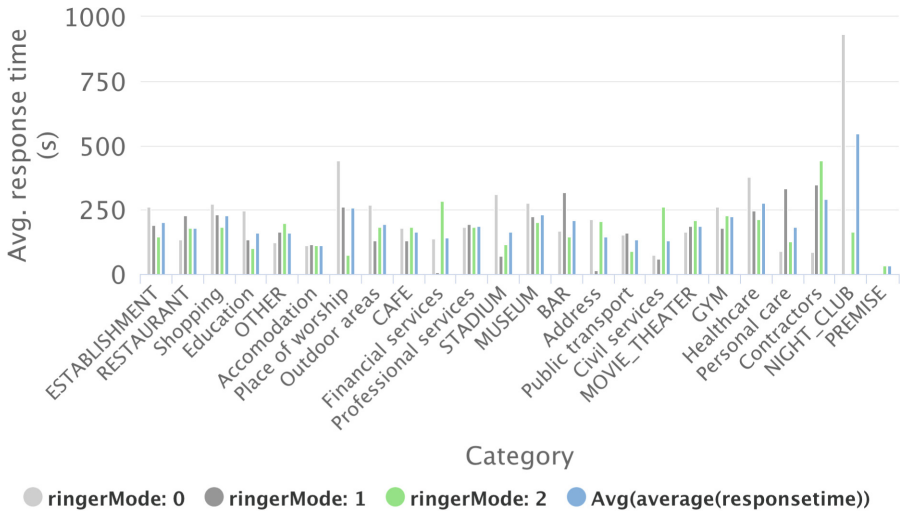


Fig. 4. Average response time per category and device ringer mode (0 = silent, 1 = vibrate only, 2 = all modalities)

4 Predicting User Location Types

4.1 Algorithms and Parameter Selection

For our analysis, we used decision trees to perform multinomial classification on the prediction target (place type), since they have been shown to demonstrate comparable performance to other methods [5]. To obtain an estimate of good parameters to use, we employed an evolutionary algorithm search process on a small hold-out dataset. The final parameters used for the algorithm are Maximal depth: 23, Minimal gain: 0.013, Minimal leaf size: 2, Minimal split size: 4. Throughout the analysis reported in the following sections, we used a 10-fold cross-validation approach. As per other studies, we note that there is an imbalance in the frequency of location categories (Table 3), hence for performance we adopt the F1-score (macro-averaged), which is more appropriate for imbalanced datasets, compared to the accuracy measure usually encountered in previous literature.

4.2 Decision Tree Modelling Performance

As a starting point, we apply the decision tree classification algorithm to the entire dataset. To clarify the process further, the classification algorithm is fed with all features as shown in Table 2, and returns the predicted place category. We assume the user’s location is the same as each venue’s reported coordinates. Therefore, given the user’s notification handling behaviour, their location, and the device state, we attempt to predict the type of venue that they are currently at. As seen in Fig. 5, the classification performance is quite good for most categories (F-score macro μ . 82.9%, σ = 12.6%). During analysis, we noted that there is some discrepancy in the confidence reported for the most likely current user place, across the place categories (Fig. 6). For this reason, we decided to repeat the analysis in multiple steps, each time limiting the dataset to contain only notifications reported where the most likely current user place was reported above a certain confidence threshold $T \in [0, 0.1, ..0.9]$. The results are shown in Fig. 7. We note that the average F-score is not majorly affected by the reduction of the dataset, however the best nominal performance is achieved when considering venues reported with a confidence threshold $T \geq 0.7$ (μ = 84.6%, σ = 13.51%, dataset size = 13,558 entries).

4.3 Modelling with Inaccurate User Coordinates

In the preceding analysis, we assumed that a user’s current coordinates are the same as those corresponding to places reported by Google’s API. Of course, it would be rare that the user’s actual coordinates would be precisely the same as those that match a specific venue, especially for venues that cover a large area (e.g. outdoor parks). To overcome this limitation, we proceeded to modify the user’s coordinates by adding random noise to the known place coordinates (latitude and longitude). This noise was applied to each coordinate component

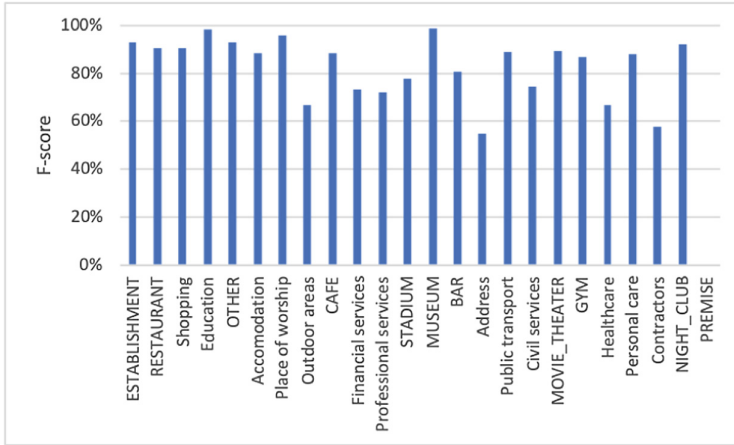


Fig. 5. Average F-score using decision trees, all notifications

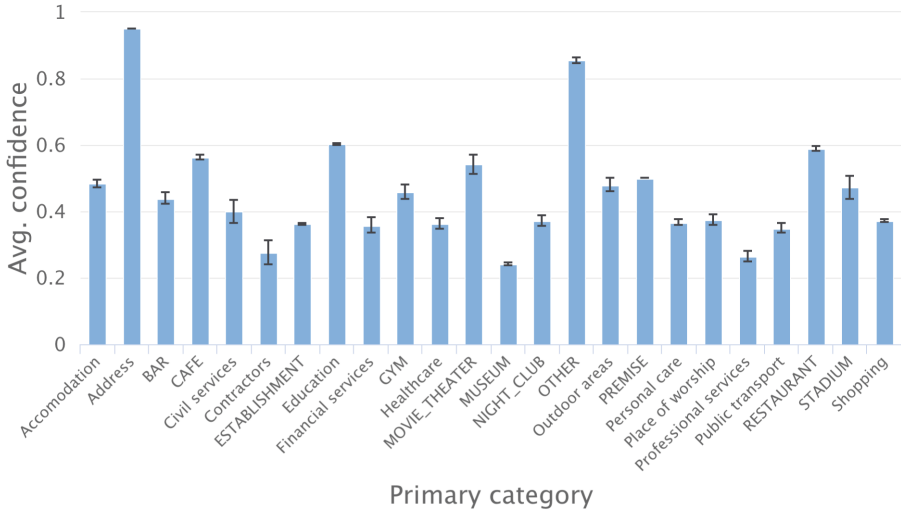
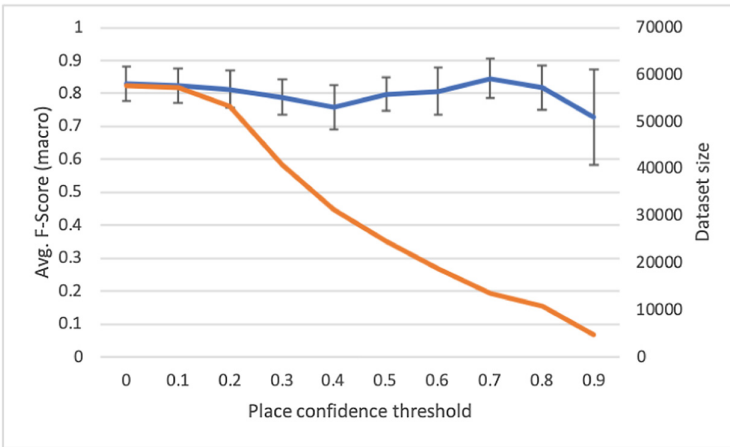


Fig. 6. Average confidence of most likely user place, all notifications, error bars at 95% c.i.

individually, following a Gaussian distribution with a standard deviation set by us. The noise standard deviation was calculated using the formula $n \times 10^{-x}$ and was applied to each coordinate component (latitude and longitude), therefore the resulting random coordinates would fall within a certain circular range of a specific venue. An example of how this process generates the random user coordinates within a gaussian distance distribution of a specific venue is shown in Table 4. Distance is calculated using the Haversine formula.

Table 4. Sample random coordinate range generation

Noise σ	Lat	Lng	Dist. at 1σ (m)
0 (Place coords.)	38.2836678	21.7889705	0
1.0×10^{-6}	38.28370608	21.78899229	4.7
2.0×10^{-6}	38.28374437	21.78901408	9.3
3.0×10^{-6}	38.28378265	21.78903587	14.0
4.0×10^{-6}	38.28382093	21.78905766	18.6
5.0×10^{-6}	38.28385922	21.78907944	23.3

**Fig. 7.** Average F-score using decision trees (error bars at 95%c.i.)

To assess the effect of imprecise user coordinates, we repeated the analysis for each value of $n \in [1, 2, \dots, 9]$, limiting the dataset to locations with a confidence threshold $T \geq 0.7$, since this achieved the best nominal performance in the preceding analysis. As can be seen in Fig. 8a, the algorithm remains quite robust when adding noise to the decimal coordinates with a $\sigma \leq 9 \times 10^{-6}$ ($\approx 42\text{m}$), after which, performance begins to deteriorate significantly. We performed also the process for a few larger distances (Fig. 8b). As expected, the performance degradation continues.

At this point, it becomes interesting to observe which categories suffer the heaviest penalty then the user coordinates are further away from the actual place coordinates. Taking the largest noise σ distance (233.1 m), we note that the categories Place of worship, Outdoor areas, Professional services, Stadium and Civil services take the worst hit between -35.46% and -55.45% reduction of their F-score, compared to the smallest σ (4.7 m). On the other hand, some categories like Shopping and Cafe only take a small penalty (-7.56% and -7.20%) respectively. The explanation for this is possibly rests in the spatial clustering of these venue types (e.g. see Fig. 9). In Fig. 9, we see that cafes are mostly clustered

together, hence we may not be able to accurately guess *exactly which* cafe a user is at, but we can be quite certain that they might be at *some* cafe, as long as their location and notification response behaviour is proximal to that captured at a nearby cafes. Although this might suggest that spatial distribution may have a significant effect on the accuracy of the classifier, it must be borne in mind that this is a very extreme scenario. Most users' location data is obtained via assisted GPS (A-GPS), which, in an urban environment, has been shown to have an accuracy of about 9m [14].

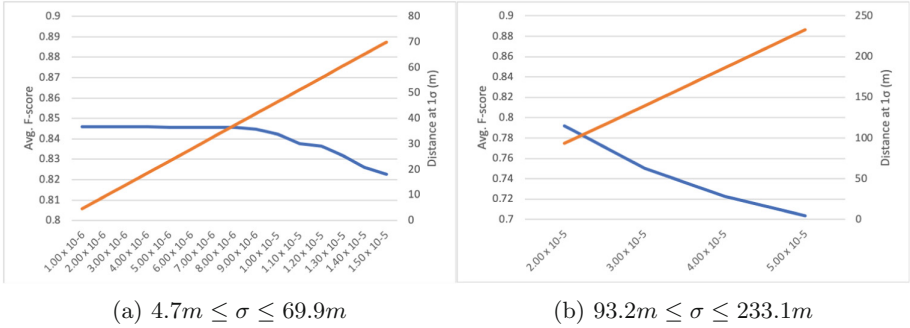


Fig. 8. Average F-score using decision trees, under random coordinate input noise

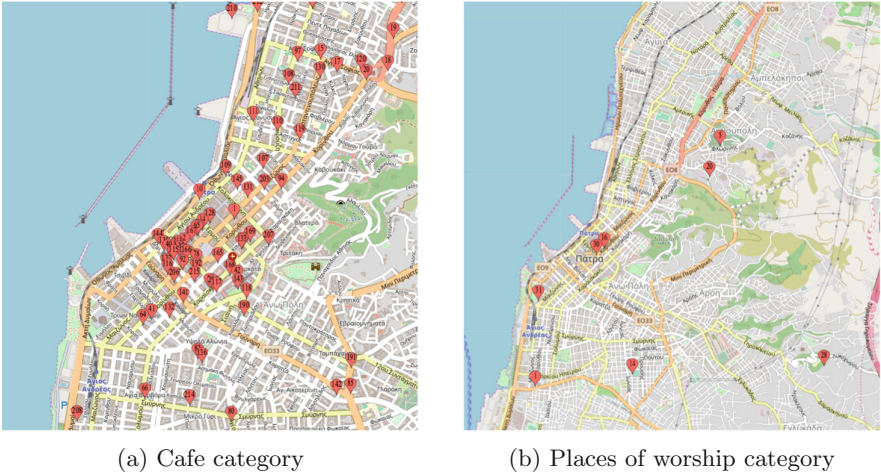


Fig. 9. Spatial distribution of places in our dataset

4.4 Effect of User Location Coordinates

In the preceding analysis, one of the input features is the user’s location. This feature is certainly obtainable from the user, but its availability depends on whether a user has enabled positioning on their device, their surroundings (indoors or outdoors) and connectivity (wi-fi, 4G, off). So far we have demonstrated that guessing the user’s current location type is possible based on their notification behaviour, device state and geographic location, even if the latter is not precisely correspondent to a known place. For the next step, we wanted to experiment without taking user position coordinates into account. The same process as in the previous analysis was repeated, limiting the dataset iteratively to contain notifications at locations above a confidence threshold T . As shown in Fig. 10 the results are much worse than in our previous analysis, showing that the prediction model depends heavily on the knowledge of the user’s coordinates, even though these do not necessarily need to correspond with great precision to the true location’s coordinates.

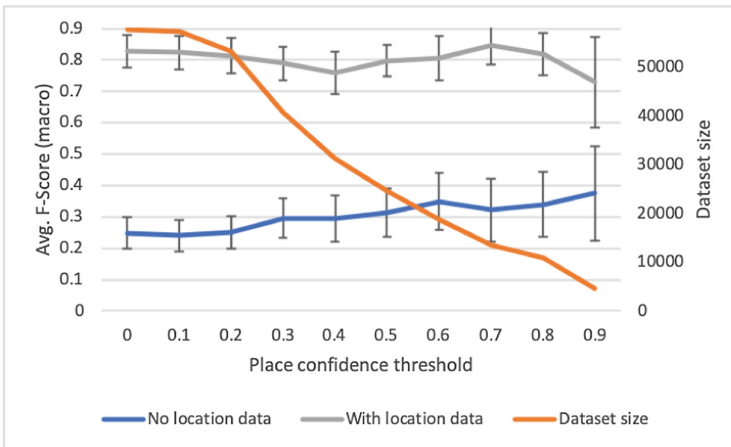


Fig. 10. Average F-score using decision trees (error bars at 95% c.i.)

4.5 Discussion

In this paper, we examined the use of notification handling behaviour as a cue for semantically labelling the user’s current location. We found that, when paired with location coordinates, the resulting models can yield useful results with high classification accuracy. Such models can be pre-trained on the cloud and then stored and ran locally on the user’s device, as part of an application or service framework, without the need for an internet connection. Further, we demonstrated that such models are robust to deviations up to 42m of user coordinates from the actual place coordinates, thereby allowing for positioning errors, or even, the deliberate obfuscation of precise user coordinates, in order to maintain privacy.

An underlying assumption in our analysis is that the user is currently positioned and has a certain non-trivial stay time at the location where the notification was received. This is likely true for most cases, as users spend more time stationary at various places, than being mobile. However, further work here could include filtering of notification events during transit times, which in our case could not be done (since we did not keep GPS logs for privacy).

For all preceding analyses, there is another underlying assumption, which is that Google’s labelling of the place categories has been used as the ground truth. However, this labelling is not necessarily correct. Since many of these places are added by users (and ostensibly curated by a few moderators), the assignment of place categories is not necessarily precise. While for some venue categories this issue might be less pronounced, categories such as “Establishment” are quite vague and therefore likely to contain many inaccuracies. As an example, our dataset contains 513 distinct places of type “Establishment”. A manual search of this list reveals that 33 of these venues would be better classified under the type “Education” (e.g. “Department of Civil Engineering”). Therefore, it must be noted that as with other studies that leverage social network data (e.g. [5]) the algorithms are tuned to predict the ground truth as reported by the location identification services, therefore introducing an inherent element of inaccuracy. In future work, it would be interesting to examine the failings of these classification algorithms, although this would require a significant investment of time and effort to obtain a reliable ground truth. However, where such algorithms fail, there might be an opportunity to exploit these failures in order to flag venues with incorrect labelling, helping thus to better curate such location datasets.

Finally, as we note different behaviours across venue categories, it would be of value to learn the reasons leading to these variances in user behaviour. However, this would be the subject of a further qualitative study. The generalisability of the findings presented here is limited to the body of the participants (students), hence the varying distribution of sample across categories. The models can be improved by mining information from other populations, to build up the number of samples across as many categories as possible. Personalised models depending on user type can then also be applied to better improve classification performance.




References

1. Akosa, J.S.: Predictive accuracy: a misleading performance measure for highly imbalanced data. In: SAS Global Forum 2017, Orlando, FL, USA, April 2017
2. Anderson, C., Hübener, I., Seipp, A.K., Ohly, S., David, K., Pejovic, V.: A survey of attention management systems in ubiquitous computing environments **2**(2), 58:1–58:27. <https://doi.org/10.1145/3214261>
3. Auda, J., Weber, D., Voit, A., Schneegass, S.: Understanding user preferences towards rule-based notification deferral. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, CHI EA 2018, pp. LBW584:1–LBW584:6. ACM. <https://doi.org/10.1145/3170427.3188688>
4. Celik, S.C., Incel, O.D.: Semantic place prediction from crowd-sensed mobile phone data **9**(6), 2109–2124. <https://doi.org/10.1007/s12652-017-0549-6>

5. Falcone, D., Mascolo, C., Comito, C., Talia, D., Crowcroft, J.: What is this place? Inferring place categories through user patterns identification in geo-tagged tweets. In: 6th International Conference on Mobile Computing, Applications and Services, pp. 10–19. <https://doi.org/10.4108/icst.mobicase.2014.257683>
6. Gu, Y., Yao, Y., Liu, W., Song, J.: We know where you are: home location identification in location-based social networks. In: 2016 25th International Conference on Computer Communication and Networks (ICCCN), pp. 1–9. <https://doi.org/10.1109/ICCCN.2016.7568598>
7. He, T., Yin, H., Chen, Z., Zhou, X., Sadiq, S., Luo, B.: A spatial-temporal topic model for the semantic annotation of POIs in LBSNs **8**(1), 12:1–12:24. <https://doi.org/10.1145/2905373>
8. Komninos, Andreas, Frengkou, Elton, Garofalakis, John: Predicting user responsiveness to smartphone notifications for edge computing. In: Kameas, Achilles, Stathis, Kostas (eds.) *AmI 2018*. LNCS, vol. 11249, pp. 3–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-03062-9_1
9. Krumm, J., Rouhana, D.: Placer: semantic place labels from diary data. In: Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp 2013, pp. 163–172. ACM (2013). <https://doi.org/10.1145/2493432.2493504>
10. Leppäkoski, H., et al.: Semantic labeling of user location context based on phone usage features. <https://www.hindawi.com/journals/misy/2017/3876906/>, <https://doi.org/10.1155/2017/3876906>
11. Saikia, P., She, J.: Effective mobile notification recommendation using social nature of locations. In: 2017 IEEE 15th International Conference on Dependable, Autonomic and Secure Computing, 15th International Conference on Pervasive Intelligence and Computing, 3rd International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), pp. 1265–1270 (2017). <https://doi.org/10.1109/DASC-PiCom-DataCom-CyberSciTec.2017.203>
12. Visuri, A., van Berkel, N., Okoshi, T., Goncalves, J., Kostakos, V.: Understanding smartphone notifications’ user interactions and content importance **128**, 72–85. <https://doi.org/10.1016/j.ijhcs.2019.03.001>
13. Wu, X., Chen, L., Lv, M., Han, M., Chen, G.: Cost-sensitive semi-supervised personalized semantic place label recognition using multi-context data **1**(3), 116:1–116:14. <https://doi.org/10.1145/3131903>
14. Zandbergen, P.A.: Accuracy of iPhone locations: a comparison of assisted GPS, WiFi and cellular positioning **13**(s1), 5–25. <https://doi.org/10.1111/j.1467-9671.2009.01152.x>
15. Zhu, Y., Zhong, E., Lu, Z., Yang, Q.: Feature engineering for semantic place prediction **9**(6), 772–783. <https://doi.org/10.1016/j.pmcj.2013.07.004>



Data-Driven Intrusion Detection for Ambient Intelligence

Ioannis Chatzigiannakis¹ , Luca Maiano¹, Panagiotis Trakadas²,
Aris Anagnostopoulos¹ , Federico Bacci¹, Panagiotis Karkazis³,
Paul G. Spirakis^{4,5} , and Theodore Zahariadis²

¹ Sapienza University of Rome, Rome, Italy
ichatz@diag.uniroma1.it, lucamaiano@gmail.com, aris@dis.uniroma1.it,
fedeb703@gmail.com

² National and Kapodistrian University of Athens, Athens, Greece
{ptrakadas,zahariad}@uoa.gr

³ University of West Attica, Aigaleo, Greece
p.karkazis@uniwa.gr

⁴ Computer Science Department, University of Liverpool, Liverpool, UK
p.spirakis@liverpool.ac.uk

⁵ Computer Engineering and Informatics Department,
Patras University, Patras, Greece

Abstract. Billions of embedded processors are being attached to everyday objects and houseware equipment to enhance daily activities and enable smart living. These embedded processors have enough processing capabilities to process sensor data to produce smart insights, and are designed to operate for months without the need of physical interventions. Despite the compelling features of Internet of Things (IoT), applied at several home-oriented use cases (e.g., lighting, security, heating, comfort), due to the lack of a physical flow of information (e.g., absence of switches and cable-based gateways), the security of such networks is impeding their rapid deployment. In this work we look into IPv6 based IoT deployments, since it is the leading standard for interconnecting the wireless devices with the Internet and we propose a data-driven anomaly detection system that operates at the transport-layer of 6LoWPAN deployments. We present a comprehensive experimental evaluation carried out using both simulated and real-world experimentation facilities that demonstrates the accuracy of our system against well-known network attacks against 6LoWPAN networks.

1 Introduction to Security Issues and Vulnerabilities in Ambient Intelligence

As IoT devices are becoming integral parts of our homes, a number of issues related to privacy and trust need to be addressed. Data confidentiality, authentication, access control within the IoT network, privacy and trust among users and devices, and the enforcement of security and privacy policies are among these

issues. However, the different standards and communication stacks involved in combination with the wide variety of embedded hardware components make traditional security countermeasures difficult to be directly applied in the IoT domain. Moreover, the high number of interconnected devices arises scalability issues [28].

The vision of the IoT has led to a competitive market for stakeholders that, in the absence of common standards, marketed proprietary and application-specific solutions, including a variety of hardware platforms, operating systems, communication protocols and data management schemes. Substantial standardisation progress has been made by different bodies (e.g., IETF CoAP, RPL, 6LoWPAN, 6Tsch, IEEE 802.15.4e, ETSI M2M, 3GPP MTC, oneM2M), providing technical solutions tailored to the resource-constrained embedded nodes, ranging from the lower to the upper OSI layers. Yet, until now, no standard has managed to attract the vast majority of the stakeholders and dominate the domain. This comes as no surprise as in most cases embedded systems are indeed application-specific and no single protocol stack can address all possible functional and non-functional requirements and specifications [13].

Clearly satisfying the security and privacy requirements of each different connectivity architecture and protocol stack is a complicated task. Moreover, the special operating conditions of IoT networks makes it even more challenging [11]. The wireless devices are more vulnerable to various attacks as their location is not known at design time and protection against tampering is very difficult due to their low cost. Therefore, it is easy to assume that the adversary can easily capture the devices, and easily read the content of their memory; thus learning the cryptographic secrets and possibly modifying their behavior. In addition, the high node-to-human ratio makes it infeasible to even consider the presence of an online trusted server that monitors and maintains individual nodes constantly. Thus techniques of pre-distribution of keys are much less effective than in traditional networks [9]. Furthermore, as sensor nodes are battery operated, security systems must reduce the energy consumption. Also, since sensor nodes have limited computing and storage capability, cryptographic algorithms and protocols that require intensive computation, communication, or storage are simply not applicable in sensor networks [3, 8]. It is too costly (in terms of computation) to authenticate using a public key and too costly (in terms of memory and computation) to store one-way chains of keys.

These constraints greatly increase the difficulty of securing IoT networks and make them more vulnerable to security threats. Still, since building secure IIoT networks is of paramount importance, the only viable solution is to combine different techniques for securing the system, i.e., implement secure routing schemes, secure aggregation, provide group key establishment methods, cryptographically encrypt messages etc.; although each single level defense mechanism is highly vulnerable, the combination of multiple attacking angles increases the overall achieved security. Towards this direction, in this paper we focus on the use of intrusion detection systems as an additional mechanism to further improve the security of IIoT networks. In contrast to the other approaches, intrusion

detection can protect from both inside and outside adversaries. Compared to existing systems for IIoT networks, our approach is more complete in the sense that it addresses all the levels of the node stack and also is more energy efficient.

1.1 Security Threats

Unlike in typical stand-alone wireless network deployments, IoT have a different purpose, in the sense that communications will not involve human interaction. This differentiates the challenges for securing the system and therefore the approaches for offering protection need to be reconsidered. There are many publications [5, 20, 24, 31] that consider some of the most significant security problems. In this work we try to summarize all the existing threats and point out the major attacks against IIoT. We call an IIoT node a *normal node* if it operates based on the system specifications. Otherwise, it is a *malicious node* or an *adversary*.

Wireless Threats. The most basic threats are due to the nature of communication that takes place over a wireless channel. Wireless communication suffer from a number of vulnerabilities:

1. **Eavesdropping.** The most easy way is to overhear the information that a node transmits or receives and then analyze the captured data and extract sensitive information without interacting with the network. These attacks are also known as *passive attacks*.
2. **Data alteration.** An adversary can cause collisions of wireless transmissions and then try to modify the message exchanged between wireless parties. These attacks are also known as *active attacks*.
3. **Identity theft.** Untethered in nature, the adversary can impersonate a legitimate user and when the original user is inactive, transmit messages without being noticed.

Routing Threats. The simplicity of many routing protocols for IIoT networks makes them an easy target for attacks. Karlof and Wagner in [21] classify the routing attacks into the following categories:

1. **Spoofed, altered, or replayed routing information.** While sending the data, the information in transit may be altered, spoofed, replayed, or destroyed. Since sensor nodes usually have only short range transmission, an attacker with high processing power and larger communication range could attack several sensors simultaneously and modify the transmitted information.
2. **Selective forwarding.** In this kind of attack a malicious node may refuse to forward every messages it gets, acting as black hole or it can forward some messages to the wrong receiver and simply drop others.
3. **Sinkhole attacks.** In the Sinkhole attack, the goal of the attacker is to attract all the traffic. Especially, in the case of a flooding based protocol

the malicious node may listen to requests for routes, and then reply to the requesting node with messages containing a bogus route with the shortest path to the requested destination.

4. **Sybil attacks.** In Sybil attack the compromised node presents itself with multiple nodes identity. This type of attack tries to degrade the usage and the efficiency of the distributed algorithms that are used. Sybil attack can be performed against distributed storage, routing, data aggregation, voting, fair resource allocation, and misbehavior detection [25].
5. **Wormholes.** Wormhole attack [12] is an attack in which the malicious node tunnels messages from one part of the network over a link, that doesn't exist normally, to another part of the network. The simplest form of the wormhole attack is to convince two nodes that they are neighbors. This attack would likely be used in combination with selective forwarding or eavesdropping.
6. **HELLO flood attacks.** This attack is based on the use by many protocols of broadcast *Hello* messages to announce themselves in the network. So, an attacker with greater range of transmission may send many *Hello* messages to a large number of nodes in a wide area of the network. These nodes are then convinced that the attacker is their neighbor. Consequently the network is left in a state of confusion.
7. **Acknowledgment.** Some IIoT network routing algorithms require link layer acknowledgments. A compromised node may exploit this by spoofing these acknowledgments, thus convincing the sender that a weak link is strong or a dead sensor is alive.

Denial of Service (DoS). This class of attacks is not concerned with the information that is transmitted. Rather, the goal of the attacker is to exhaust the resources of the network and cause it not to function properly. Wood and Stankovic [32, 33] classify several forms of DoS attacks based on the layer that the attack uses. Some of these were already mentioned so we will not repeat them. At the physical layer the attacks take the form of jamming and tampering [24]. Jamming is done by interfering with the radio frequencies nodes are using. Tampering refers to the physical altering or even damaging of the nodes. An attacker can damage and replace a node, for example, by stealing or replacing information or cryptographic keys. At the link layer the attacker can generate collisions and exhaustion may be caused by protocols that attempt retransmission repeatedly, even when triggered by an unusual and suspicious collision.

1.2 Security Solutions

Several attempts have been made in the past years towards preventing intrusions over IoT networks based on alternative encryption and authentication techniques. Some solutions focus on the validation of the integrity of message exchanges in order to securely route information across the wireless medium [30, 36]. Others propose lightweight, distributed secure group communication primitives that operate on-top of the networking layer to protect data and

to cope with potential compromises [10, 14], while others work at the application layer and propose energy-efficient encryption methods to guarantee the confidentiality of the information exchanged [6]. However, these approaches cannot completely prevent malicious users from intruding the network and tamper with the operation of the network [7]. For example, the above techniques may not be fully protected against compromised nodes which participate in the network and already have the shared cryptographic keys [35].

In the context of the IP-based IoT solutions, consideration of TCP/IP security protocols is important as these protocols are designed to fit the IP network concept and technology. While a wide range of specialized as well as general-purpose key exchange and security solutions exist for the Internet domain, the 6LoWPAN and CoRE IETF working groups look into IKEv2/IPsec [22], TLS/SSL [27], DTLS [26], HIP [23], PANA [16], and EAP [1] as candidate solutions of IIoT, we focus on the discussion of in this paper. Application layer solutions such as SSH [34] also exist, however, these are currently not considered.

1.3 Anomaly Detection for Internet of Things

In an attempt to improve the overall security of networks, a number of Intrusion detection systems (IDS) have been proposed [6]. Such IDS architectures are classified into two basic categories depending on the *data collection mechanism*:

- **Host-based.** The IDS consults several type of log files (kernel, system, application, etc.) and compares the logs against an internal database of common signatures for known attacks.
- **Network-based.** The IDS is scanning network packets, auditing packet information, and logging any suspicious packets.

IDS architectures can be further classified based on the *detection technique*:

- **Signature-based.** IDS centers on finding an occurrence of predefined signatures or behavior that matches a previously known malicious action or indicates an intrusion.
- **Anomaly-based.** IDS checks for any behaviors that fall outside the predefined or accepted model of behavior.
- **Specification-based.** IDS defines a set of constrains that are indicative of a program's or protocol's correct operation.

Furthermore, *IDS architectures specific to wireless ad-hoc networks* are further divided into three categories:

- **Stand-alone.** Each node operates as a independent IDS and is responsible for detecting attacks only for itself. Such an IDS does not share any information or cooperate with other systems. This architecture implies that all the nodes of the network are capable of running an IDS.
- **Distributed and Cooperative.** All nodes are running their own IDS, but they also cooperate in order to create a global intrusion detection mechanism.

- **Hierarchical.** The network is divided into clusters with cluster-head nodes. These nodes are responsible for routing within the cluster and accept all the accusation messages from the other cluster-members indicating something malicious. Additionally, the cluster-head nodes may also detect attacks against the other cluster-head nodes of the network, as they constitute the backbone of the routing infrastructure.

2 Materials and Method

Existing security solutions cannot completely prevent malicious users from intruding the network and tamper with the operation of the network. In this paper, we look into intrusion detection as a second line of defence to protect IoT once an intrusion is detected by activating certain actions to minimize damages, gather evidence for the prosecution, and even launch counter-attacks. We propose an *Anomaly-based* intrusion detection system (IDS) that operates at the transport-layer level on top of 6LoWPAN networks. Assuming an IPv6 enabled IoT, a solution that operates at transport-layer is generic enough to be applied also in combination with closed systems that implement proprietary hardware and custom network protocols. Our solution does not require any particular implementation of 6LoWPAN or RPL, or any additional encryption primitive.

Our IDS uses the ICMPv6 control messages defined within 6LoWPAN to *passively* collect high-level information for the state of the network, such as the round-trip-time, hop distance and packet loss. The IDS periodically transmits a sequence of ICMPv6 control messages to each node of the network based on the public IP addresses. Then it checks for any behaviours that fall outside the predefined, accepted model of behaviour.

In contrast to existing general purpose security systems, in IoT, it is simply infeasible to expect a network administrator to be aware of the full range of potentially relevant possibilities and be able to pull them together manually. For this reason, we completely avoid defining a series of hard-coded alarm limits associated with assumed “steady states” for the network, known to send a large number of false alarms. Instead, we use a combination of machine-learning and statistical analysis to address the specific problem of anomaly detection.

We assume that there is an initial period of time where the network administrator can collect sufficient data regarding the operation of the network. Then periodically we examine the operation of the network by collecting additional statistics and try to identify observations which differ significantly from the majority of the initial data.

Given a set of traces collected: a *network-level anomaly detector* characterizes if the network is affected by at least 1 malicious node; a *class-sensitive network-level anomaly detector* characterizes if the network is affected by at least 1 malicious node of a specific attack class; a *node-level anomaly detector* characterizes each node of the network if it is a malicious node or not; and a *class-sensitive node-level anomaly detector* characterizes each node of the network if it is a malicious node of a specific attack class. The performance of an anomaly detector

is characterized by the *precision*, that is the ability of the detector not to label as positive a sample that is negative and is measured by the ratio of correctly labelled as L to all items actually labelled as L; the *accuracy*, representing the set of labels detected that exactly match the corresponding set of actual labels; the *sensitivity*, the ability of the detector to find all the positive samples and is measured by the ratio of items correctly labelled as L to all items that belong to label L; and the *overhead*, the number of packets that constitute the trace.

2.1 Experimental Scenario

The target of our experiments aim at providing an automated way of recognizing intrusions in a network. Given an RPL DODAG, we generate ICMP packets for each node of the network. Figure 1 depicts our proposed methodology.

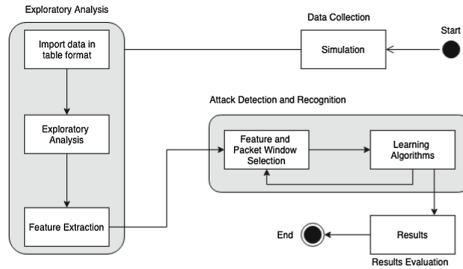


Fig. 1. Methodology flow diagram.

The experiments are based on simulated networks generated by the Cooja tool, the standard Contiki network simulator. We conducted several experiments on different grid schemes (i.e., 3×3 , 4×4 and 5×5 grids) as well as random topologies (of 9, 16, 25 nodes). For each node, we collected 200 ICMP packets (sending ping messages from the root node to each other node of the network in every 5 s), the round trip time (RTT) of each of these messages and the distance of the node from the RPL DODAG root. We parse raw data collected from each experiment to extract relevant features in a table structure. This is helpful to gather statistical information from the dataset. From the analysis of this set of data, we expected to retrieve enough information to learn the usual behaviour of the network.

Given the generated network topologies, we substitute some of the nodes of the network with malicious nodes realizing either Black Hole or Grey Hole attacks. The number of malicious nodes is significantly fewer than the number of benign nodes. This is typical in anomaly-detection problems. Since all nodes in the experiment run RPL protocol, we expect that an attacked node will affect the transmissions of its neighbours. Even more, we realized that the malicious nodes usually have more or less important effects on the entire network. Thus all nodes belonging to an attacked network have been labelled as attacked.

After this first exploratory analysis of the data, we select a set of features based on the statistics of the ICMP packets collected over a non-overlapping window of ICMP packets, to train and evaluate the learning algorithms. ICMP packets window is a subset obtained dividing the 200 ICMP messages received by a node by a fixed number in the interval [12, 24, 48, 100, 200] and extract features on that number of ICMP packets. Finally, we run the learning algorithms on the selected dataset and collect results.

2.2 Methods for Anomaly Detection

K-Nearest Neighbors (KNN). KNN algorithm is a non-parametric method used for classification and regression [2]. The input consists of the k closest training examples in the feature space. An object is classified by a plurality vote of its neighbours, with the object being assigned to the class that is most common among its k nearest neighbours (k is a positive integer, typically small).

Random Forest (RF). Random Forest (RF) algorithms employ a technique known as bagging, whereby data instances are resampled multiple times to produce multiple training subsets from the training data [4]. Decision trees are then created from each training subset until ensembles of trees have been created. Each tree then casts a unit vote for the outcome of an incoming data instance class label. RF is flexible, with constrained computational resources required.

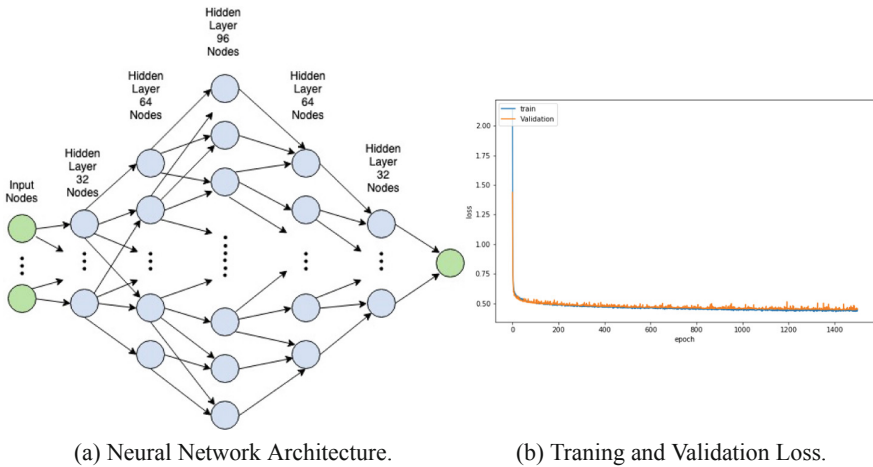


Fig. 2. Neural Network model for attack detection.

Deep Neural Network (DNN). Artificial neural networks (ANN) are computing systems that are inspired by the biological neural networks that constitute human brains (Fig. 2). Each neuron performs some calculation and outputs a value that is then spread through all its outgoing connections as input into other units [19]. Connections are characterized by weights that correspond to the importance of the link between two neurons. The computation performed by a unit is separated into two stages: the aggregation and the activation functions. The aggregation function calculates the sum of the inputs received by the unit through all its incoming connections. The resulting value is then fed into the activation function. Neurons are organized in levels. Layers between input and output layers are called the hidden layers. When a new input is given, information passes across layers until the output, where classification happens.

Support Vector Machines (SVM). The basic idea of SVM for time-series approximation is mapping the data into a high-dimensional feature space by a nonlinear mapping and then performing a linear regression in the feature space [29]. The nonlinear mapping can be efficiently computed through a kernel function, without iterating over all the corresponding data points. Given the kernel function, the SVM learner tries to find a hyperplane that separates positive from negative data points and at the same time maximizes the separation (margin) between them. This method is known to be resilient to overfitting and to have good generalization performance, due to the max-margin criterion used during optimization. Furthermore, the SVM is guaranteed to converge to a global optimum due to the corresponding convex optimization formulation.

2.3 Prediction Performance Measure

The performance of the classifiers is based on calculating the precision, accuracy, recall and F1 Score for each class and then computing their unweighted mean. This approach does not take label imbalance into account, e.g., when 95% of items are labelled Normal and 5% are labelled Under-Attack, if all the items are labelled as Normal, the accuracy would be 95% but all items from label Under-Attack would be misclassified. For this reason we made sure that there is a good balance across the different labels considered during both training and validation. The four metrics used are defined as follows:

Precision (or Positive Predictive Value (PPV)) – represents the ratio of items correctly labelled as L to all items actually labelled as L, that is, the ratio $\frac{TP}{(TP+FP)}$, where TP is the number of true positives and FP is the number of false positives. The precision is intuitively the ability of the classifier not to label as positive a sample that is negative.

Accuracy (or Proportion Correct) – represents the set of labels predicted that exactly match the corresponding set of actual labels.

Recall (or Sensitivity or True Positive Rate or Probability of Detection (PD) or Detection Rate) – represents the ratio of items correctly labelled as L to

all items that belong to label L , that is, the ratio $\frac{TP}{(TP+FN)}$, where TP is the number of true positives and FN is the number of false negatives. The recall is intuitively the ability of the classifier to find all the positive samples.

F1 score, or balanced F-score or F-measure represents a weighted average of the precision and recall, where 1 is the best score and 0 the worst score. The relative contribution of precision and recall to the F1 score are equal.

The formula for the F1 score is: $\frac{2 \times (\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})}$.

AUC (or Area Under the ROC curve) – ROC is a probability curve and AUC represents degree or measure of separability. It tells how much model is capable of distinguishing between classes.

2.4 Implementation Details

The learning algorithms are implemented using Python 3.7.1 in combination with Keras 2.2.4, Pandas 0.24.1, Scikit 0.20.3 and Numpy 1.15.4. All the experiments have been performed on a Macbook Pro 2017 - 2, 3 GHz Intel Core i5, 8 GB 2133 MHz LPDDR3. We release all the dataset and the code used for the experiments on a Github repository (<https://github.com/ichatz/iot-netprofiler>).

All mentioned algorithms have been extensively used by the state of the art solutions in order to solve anomaly-detection problems [15]. We used the following operating parameters:

1. **K-Nearest Neighbor (KNN)** - using $K = 3$.
2. **Random Forests Classifier** - building 100 estimators.
3. **Support Vector Machines (SVM)** - using a linear kernel.
4. **Deep Neural Network Classifier** - using 5 hidden layers of 32, 64, 96, 64, 32 neurons respectively. Dropout and L2 regularization are applied to reduce overfitting. We use the cross-entropy loss function to evaluate training and validation losses.
5. **K-Means** - with K equal to the number of classes. In this case, data are pre-processed using a PCA.

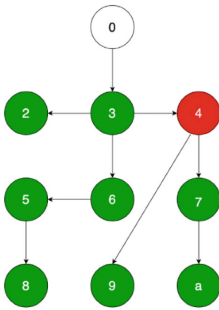
3 Results

3.1 Exploratory Analysis

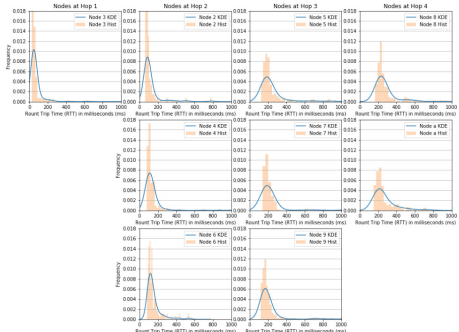
We start with by conducting a first exploratory analysis of the network statistics. Figure 3 depicts the distribution of round-trip-time (RTT) for each node, given the hop-distance from the root, for a topology involving 9 nodes in a 3×3 grid. Similar results are acquired for larger topologies. We keep in mind that the RPL DODAG is designed to have a different configuration each time an RPL instance on the network is running, thus, even with the same topology, the network could have a different structure. However, regardless of the RPL DODAG operating parameters, by simply observing the basic network statistics, it is clear that the distribution of the RTT at each hop-distance is more or less the same. Given the

distribution of the RTT of a node at certain hop-distance from the root, we can usually understand if the node (or one of its neighbours) has been attacked. In fact, depending on which kind of attack has been performed, it is easy to notice that the distribution of a certain number of nodes is affected. In particular, we can distinguish between the following cases:

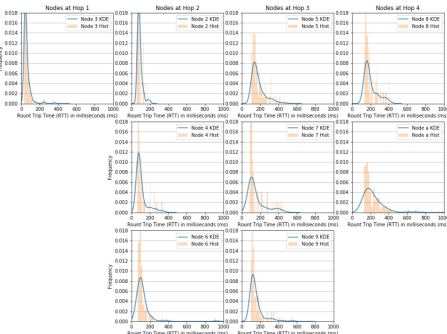
- if the network has been attacked with a Black Hole attack at a non-leaf node N , this attack also affects the children nodes of N . These nodes either remain unavailable for the entire experiment, or after a given period of time they manage to identify an alternative path connecting them with the root node;
- if a Gray Hole attack has been performed, we can still observe an effect on variance and latency values in some nodes.



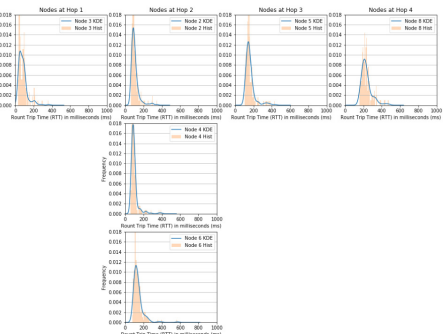
(a) Network Topology.



(b) Normal behaviour.



(c) Grey Hole Attack.



(d) Black Hole Attack.

Fig. 3. RTT distribution of 3 × 3 topology.

Since the RTT varies over time, it is important to study if there is an *Auto-correlation*, highlighting in this way the similarity between observations as a

function of the time lag between them. Studying *autocorrelation* can also identify possible *seasonability* in the measurements, i.e., periodic fluctuations. Assuming that the distribution of each variable fits a Gaussian distribution, we can use the *Pearson’s correlation coefficient* to summarize the correlation between the variables. Pearson’s correlation coefficient is a number between -1 and 1 that describes a negative or positive correlation, respectively. A value of zero indicates no correlation. We can calculate the correlation for time-series observations with previous time steps, called lags. Because the correlation of the time series observations is calculated with values of the same series at previous times, this is called a serial correlation, or an autocorrelation. A plot of the autocorrelation of a time series by lag is called the AutoCorrelation Function, or the acronym ACF.

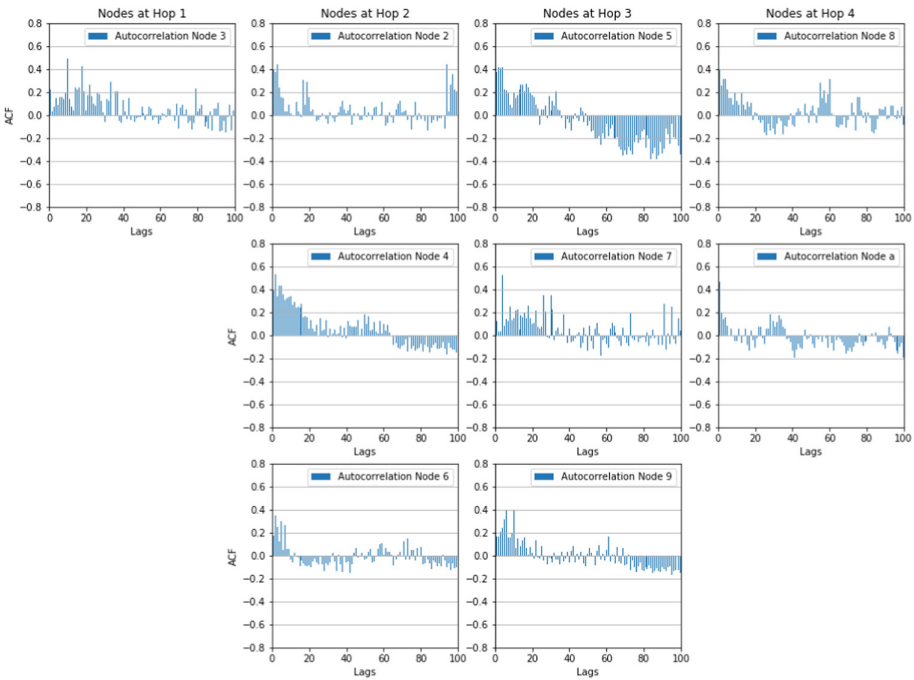


Fig. 4. Autocorrelation of a 3×3 grid experiment.

The results are included in Fig. 4, obtained above do not show particular trends, suggesting that the RTT time series are stationary. In order to make strong assumptions about our data, we computed *Advanced Dickey-Fuller test* [17]. This is a statistical test called a unit root test. The intuition behind a unit root test is that it determines how strongly a time series is defined by a trend. It uses an autoregressive model and optimizes an information criterion across multiple different lag values. The null hypothesis of the test is that the

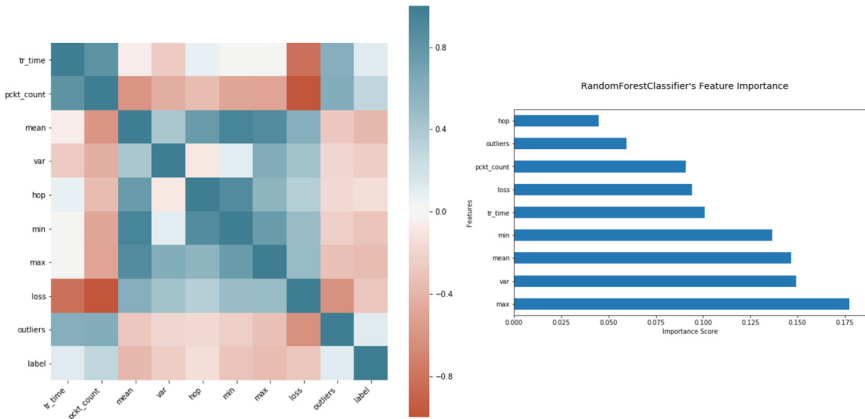
time series can be represented by a unit root, that it is not stationary (has some time-dependent structure). The alternate hypothesis (rejecting the null hypothesis) is that the time series is stationary.

1. Null Hypothesis (H0): If failed to be rejected, it suggests the time series has a unit root, meaning it is non-stationary. It has some time-dependent structure.
2. Alternate Hypothesis (H1): The null hypothesis is rejected; it suggests the time series does not have a unit root, meaning it is stationary. It does not have a time-dependent structure.

Results from Advanced Dickey-Fuller test confirms that the hypothesis that RTTs time series is stationary, at least for more than 79% of nodes. Starting from these results, a *log transform* has been used to flatten out non-stationary nodes back to a linear relationship. This could help learning algorithms; in fact, stationary time series can be easier to model.

3.2 Features Extraction and Selection

We now move one to studying the importance of different features in terms of the learning rate of machine learning and deep learning algorithms. Over a fixed window of packets, we consider the mean, variance, min, and max values of RTT, the number of **outliers** on the RTT value, the hop distance of the node from the root, and the number of lost packets. Remark that since data resulting from different networks could have different mean and variance due to their different network topologies, we apply *feature normalization*.



(a) Correlation Matrix.

(b) Random Forest Classifier Features Importance.

Fig. 5. Features correlation and importance.

Some features could be more important than others, so if we select the right set of features, this could help to improve the results given by a learning classifier. In order to select the best set of features, we calculate the correlation matrix of features. We use the Pearson Coefficient that assigns a value within 1 (i.e. positive correlation) and -1 (negative correlation). Following the approach proposed by Furkan Yusuf Yavuz, Devrim Ünal and Ensar Gul in [18], we use *Random Forest Classifier* to iteratively select the most relevant features. In fact, as they suggest, if feature importance is high, it dilutes the effect of the others and may cause overfitting, while less important could slow down (or even deviate) the learning process (Fig. 5).

3.3 Experimental Results

We now examine the performance of different machine learning algorithms to accurately detect attacks in an IoT network. We use two different experimental scenarios. In the first one, we want to detect if a network has been attacked or not. In the second scenario, we try to recognize what kind of attack has been performed if any. In each scenario, we train the learning algorithms with the following approach:

1. Select a window of N ICMP packets for each node ($N \in [12, 24, 48, 100, 200]$).
2. Select a subset of features.
3. Split the dataset using 80% of the data for training and 20% for testing.
4. Train the algorithm using 5-fold cross-validation. Therefore, split the training set to and iteratively evaluate the results on 20% of this set.
5. Finally, we test the algorithm on the test set.

The results suggest that we can accurately detect if an attack has been performed. Attacks are accurately detected by KNN, Random Forests and Deep Neural Network with an accuracy of 80%. Due to the observed stationarity, varying the size of the windows does not affect the results. Even with a window size of 12 packets, which corresponds to 1 min, we can accurately detect an attack in the network. Figure 6 shows the results of this experiment for attack detection.

model	n_classes	accuracy	precision	recall	f1-score	auc roc
knn	2	0.803406	0.797810	0.794549	0.796010	0.794549
random forest	2	0.829721	0.829888	0.815789	0.820818	0.815789
svm	2	0.752322	0.755062	0.726316	0.731972	0.726316
neural network	2	0.823529	0.847324	0.795865	0.806430	0.795865
kmeans	2	0.375891	0.383400	0.382168	0.375668	0.382168

Fig. 6. Attack detection results.

We also examine the classification of specific type of attack. Here KNN and Random Forests achieve the highest accuracy in the range of 71% and 75% accuracy. Figure 7 indicates the different values achieved for each different method considered.

model	n_classes	accuracy	precision	recall	f1-score	auc roc
knn	3	0.712963	0.698080	0.708799	0.701690	0.777039
random forest	3	0.750000	0.759638	0.727007	0.740593	0.791564
svm	3	0.616667	0.686075	0.503812	0.514424	0.628717
neural network	3	0.505556	0.168519	0.333333	0.223862	0.500000
kmeans	3	0.439585	0.433131	0.381860	0.376426	0.538705

Fig. 7. Attack recognition results.

4 Conclusions and Future Work

Our study presents an approach which can detect routing attacks based on a simple analysis of ICMP packets. Packet-drop attacks (black-hole attack and grey-hole attack) are successfully detected by our proposed attack detection models. Our methodology proposes a simple and efficient solution to detect if such attacks have been performed. Experimental results suggest that even kind of attack could be detected with high accuracy if learning models are trained with a sufficiently wide range of data samples. The lack of open-source datasets and the difficulty in collecting this type of data surely makes it difficult to train learning algorithms efficiently. As future work, we intend to extend the dataset with additional attack types and scenarios, introducing a different rate of malicious and normal nodes as well as a larger number of nodes. We also plan to test more sophisticated Neural Network architectures like LSTM, that could help to achieve better results on a larger dataset.

References

1. Aboba, B., Blunk, L., Vollbrecht, J., Carlson, J., Levkowitz, H.: Extensible authentication protocol (EAP) (2005)
2. Altman, N.S.: An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **46**(3), 175–185 (1992)
3. Boukerche, A., Chatzigiannakis, I., Nikolettseas, S.: Power-efficient data propagation protocols for wireless sensor networks. *Simulation* **81**(6), 399–411 (2005)
4. Breiman, L.: Bagging predictors. *Mach. Learn.* **24**(2), 123–140 (1996)
5. Butun, I., Morgera, S.D., Sankar, R.: A survey of intrusion detection systems in wireless sensor networks. *IEEE Commun. Surv. Tutorials* **16**(1), 266–282 (2014)
6. Chatzigiannakis, I., Pyrgelis, A., Spirakis, P.G., Stamatou, Y.C.: Elliptic curve based zero knowledge proofs and their applicability on resource constrained devices. In: 2011 IEEE Eighth International Conference on Mobile Ad-Hoc and Sensor Systems, pp. 715–720, October 2011

7. Chatzigiannakis, I., Strikos, A.: A decentralized intrusion detection system for increasing security of wireless sensor networks. In: 2007 IEEE Conference on Emerging Technologies and Factory Automation (EFTA 2007), pp. 1408–1411, September 2007
8. Chatzigiannakis, I., Kinalis, A., Nikolettseas, S.: An adaptive power conservation scheme for heterogeneous wireless sensor networks with node redeployment. In: Proceedings of the Seventeenth Annual ACM Symposium on Parallelism in Algorithms and Architectures, pp. 96–105. ACM (2005)
9. Chatzigiannakis, I., Konstantinou, E., Liagkou, V., Spirakis, P.: Design, analysis and performance evaluation of group key establishment in wireless sensor networks. *Electron. Notes Theor. Comput. Sci.* **171**(1), 17–31 (2007)
10. Chatzigiannakis, I., Konstantinou, E., Liagkou, V., Spirakis, P.: Design, analysis and performance evaluation of group key establishment in wireless sensor networks. *Electron. Notes Theor. Comput. Sci.* **171**(1), 17–31 (2007). Proceedings of the Second Workshop on Cryptography for Ad-hoc Networks (WCAN 2006)
11. Chatzigiannakis, I., Mylonas, G., Vitaletti, A.: Urban pervasive applications: challenges, scenarios and case studies. *Comput. Sci. Rev.* **5**(1), 103–118 (2011)
12. Hu, Y., Perrig, A., Johnson, D.B.: Wormhole detection in wireless ad hoc networks. In: Ninth International Conference on Network protocol (ICNP), vol. 1 (2002)
13. Dimitrios, A., Vasileios, G., Dimitrios, G., Ioannis, C.: Employing internet of things technologies for building automation. In: Proceedings of 2012 IEEE 17th International Conference on Emerging Technologies & Factory Automation (ETFA 2012), pp. 1–8. IEEE (2012)
14. Wenliang, D., Deng, J., Han, Y.S., Varshney, P.K., Katz, J., Khalili, A.: A pairwise key predistribution scheme for wireless sensor networks. *ACM Trans. Inf. Syst. Secur. (TISSEC)* **8**(2), 228–258 (2005)
15. Hassan, S.A., Hussain, F., Hussain, R., Hossain, E.: Machine learning in IoT security: current solutions and future challenges. [arXiv:1904.05735v1](https://arxiv.org/abs/1904.05735v1) (2019)
16. Forsberg, D., Ohba, Y., Patil, B., Tschofenig, H., Yegin, A.: Protocol for carrying authentication for network access (PANA) (2008)
17. Fuller, W.A.: Introduction to Statistical Time Series, 2nd edn. Wiley, Hoboken (1995)
18. ÜNAL, D., GÜL, E., YAVUZ, F.Y.: Deep learning for detection of routing attacks in the internet of things. *Int. J. Comput. Intell. Syst.* **12**(1), 39–58 (2018)
19. Gamboa, J.C.B.: Deep learning for time-series analysis. [arXivpreprint arXiv:1701.01887](https://arxiv.org/abs/1701.01887) (2017)
20. Heer, T., Garcia-Morchon, O., Hummen, R., Keoh, S.L., Kumar, S.S., Wehrle, K.: Security challenges in the IP-based internet of things. *Wirel. Pers. Commun.* **61**(3), 527–542 (2011)
21. Karlof, C., Wagner, D.: Secure routing in wireless sensor networks: attacks and countermeasures. In: Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications, pp. 113–127. IEEE (2003)
22. Kaufman, C.: Internet key exchange (IKEv2) protocol (2005)
23. Moskowitz, R., Nikander, P., Jokela, P., Henderson, T.: Host identity protocol version 2 (HIPv2) (2015)
24. Mpiziopoulos, A., Gavalas, D., Konstantopoulos, C., Pantziou, G.: A survey on jamming attacks and countermeasures in wsns. *IEEE Commun. Surv. Tutorials* **11**(4), 42–56 (2009)
25. Newsome, J., Shi, E., Song, D., Perrig, A.: The sybil attack in sensor networks: analysis & defenses. In: Third International Symposium on Information Processing in Sensor Networks, IPSN 2004, pp. 259–268. IEEE (2004)

26. Phelan, T.: Datagram transport layer security (DTLS) over the datagram congestion control protocol (DCCP) (2008)
27. Rescorla, E.: The transport layer security (TLS) protocol version 1.3 (2018)
28. Sadeghi, A.-R., Wachsmann, C., Waidner, M.: Security and privacy challenges in industrial internet of things. In: 2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC), pp. 1–6. IEEE (2015)
29. Vapnik, V.: The Nature of Statistical Learning Theory. Springer, Berlin (2013). <https://doi.org/10.1007/978-1-4757-3264-1>
30. Velivasaki, T.-H.N., Karkazis, P., Zahariadis, T.V., Trakadas, P.T., Capsalis, C.N.: Trust-aware and link-reliable routing metric composition for wireless sensor networks. *Trans. Emerg. Telecommun. Technol.* **25**(5), 539–554 (2014)
31. Wallgren, L., Raza, S., Voigt, T.: Routing attacks and countermeasures in the RPL-based internet of things. *Int. J. Distrib. Sens. Netw.* **9**(8), 794326 (2013)
32. Wood, A.D., Stankovic, J.A.: Denial of service in sensor networks. *Computer* **35**(10), 54–62 (2002)
33. Wood, A.D., Stankovic, J.A.: A taxonomy for denial-of-service attacks in wireless sensor networks. In: *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*, pp. 739–763 (2004)
34. Ylonen, T., Lonvick, C.: The secure shell (SSH) protocol architecture (2006)
35. Zhang, Y., Lee, W., Huang, Y.-A.: Intrusion detection techniques for mobile wireless networks. *Wirel. Netw.* **9**(5), 545–556 (2003)
36. Zhou, L., Haas, Z.J.: Securing ad hoc networks. *IEEE Netw.* **13**(6), 24–30 (1999)



Spoken Language Identification Using ConvNets

Sarthak¹, Shikhar Shukla²(✉), and Govind Mittal³

¹ Analytics Quotient, Bangalore, India

sarthak.sfc@gmail.com, sarthak.j@aqinsights.com

² Samsung R&D Institute India-Bangalore, Bangalore, India

shikhar.00778@gmail.com, shikhar.0077@samsung.com

³ Birla Institute of Technology and Science, Pilani, Rajasthan, India

f2014530@pilani.bits-pilani.ac.in

Abstract. Language Identification (LI) is an important first step in several speech processing systems. With a growing number of voice-based assistants, speech LI has emerged as a widely researched field. To approach the problem of identifying languages, we can either adopt an implicit approach where only the speech for a language is present or an explicit one where text is available with its corresponding transcript. This paper focuses on an implicit approach due to the absence of transcriptive data. This paper benchmarks existing models and proposes a new attention based model for language identification which uses log-Mel spectrogram images as input. We also present the effectiveness of raw waveforms as features to neural network models for LI tasks. For training and evaluation of models, we classified six languages (English, French, German, Spanish, Russian and Italian) with an accuracy of 95.4% and four languages (English, French, German, Spanish) with an accuracy of 96.3% obtained from the VoxForge dataset. This approach can further be scaled to incorporate more languages.

Keywords: Language Identification · Raw waveform · Convolutional Neural Networks · Machine learning

1 Introduction

Language Identification (LI) is a problem which involves classifying the language being spoken by a speaker. LI systems can be used in call centers to route international calls to an operator who is fluent in that identified language [12]. In speech-based assistants, LI acts as the first step which chooses the corresponding grammar from a list of available languages for its further semantic analysis [1]. It can also be used in multi-lingual voice-controlled information retrieval systems, for example, Apple Siri and Amazon Alexa.

Over the years, studies have utilized many prosodic and acoustic features to construct machine learning models for LI systems [18]. Every language is

composed of *phonemes*, which are distinct unit of sounds in that language, such as *b* of black and *g* of green. Several prosodic and acoustic features are based on phonemes, which become the underlying features on whom the performance of the statistical model depends [5, 20]. If two languages have many overlapping phonemes, then identifying them becomes a challenging task for a classifier. For example, the word *cat* in English, *kat* in Dutch, *katze* in German have different consonants but when used in a speech they all would sound quite similar.

Due to such drawbacks several studies have switched over to using Deep Neural Networks (DNNs) to harness their novel auto-extraction techniques [1, 19]. This work follows an implicit approach for identifying six languages with overlapping phonemes on the VoxForge [23] dataset and achieves 95.4% overall accuracy.

In previous studies [1, 17, 19], authors use log-Mel spectrum of a raw audio as inputs to their models. One of our contributions is to enhance the performance of this approach by utilising recent techniques like Mixup augmentation of inputs and exploring the effectiveness of *Attention* mechanism in enhancing performance of neural network. As log-Mel spectrum needs to be computed for each raw audio input and processing time for generating log-Mel spectrum increases linearly with length of audio, this acts as a bottleneck for these models. Hence, we propose the use of raw audio waveforms as inputs to deep neural network which boosts performance by avoiding additional overhead of computing log-Mel spectrum for each audio. Our 1D-ConvNet architecture auto-extracts and classifies features from this raw audio input.

The structure of the work is as follows. In Sect. 2 we discuss about the previous related studies in this field. The model architecture for both the raw waveforms and log-Mel spectrogram images is discussed in Sect. 3 along with the discussion on hyperparameter space exploration. In Sect. 4 we present the experimental results. Finally, in Sect. 5 we discuss the conclusions drawn from the experiment and future work.

2 Related Work

Extraction of language dependent features like prosody and phonemes was a popular approach to classify spoken languages [6, 16, 29]. Following their success in speaker verification systems, i-vectors have also been used as features in various classification networks. These approaches required significant domain knowledge [4, 16]. Nowadays most of the attempts on spoken language identification rely on neural networks for meaningful feature extraction and classification [7, 15].

Revay et al. [19] used the ResNet50 [9] architecture for classifying languages by generating the log-Mel spectra of each raw audio. The model uses a cyclic learning rate where learning rate increases and then decreases linearly. Maximum learning rate for a cycle is set by finding the optimal learning rate using *fastai* [11] library. The model classified six languages – English, French, Spanish, Russian, Italian and German – and achieving an accuracy of 89.0%.

Gazeau et al. [8] in his research showed how Neural Networks, Support Vector Machine and Hidden Markov Model (HMM) can be used to identify French,

English, Spanish and German. Dataset was prepared using voice samples from Youtube News [27] and VoxForge [23] datasets. Hidden Markov models convert speech into a sequence of vectors, was used to capture temporal features in speech. HMMs trained on VoxForge [23] dataset performed best in comparison to other models proposed by him on same VoxForge dataset. They reported an accuracy of 70.0%.

Bartz et al. [1] proposed two different hybrid Convolutional Recurrent Neural Networks for language identification. They proposed a new architecture for extracting spatial features from log-Mel spectra of raw audio using CNNs and then using RNNs for capturing temporal features to identify the language. This model achieved an accuracy of 91.0% on Youtube News Dataset [27]. In their second architecture they used the Inception-v3 [22] architecture to extract spatial features which were then used as input for bi-directional LSTMs to predict the language accurately. This model achieved an accuracy of 96.0% on four languages which were English, German, French and Spanish. They also trained their CNN model (obtained after removing RNN from CRNN model) and the Inception-v3 on their dataset. However they were not able to achieve better results achieving and reported 90% and 95% accuracies, respectively.

Kumar et al. [12] used Mel-frequency cepstral coefficients (MFCC), Perceptual linear prediction coefficients (PLP), Bark Frequency Cepstral Coefficients (BFCC) and Revised Perceptual Linear Prediction Coefficients (RPLP) as features for language identification. BFCC and RPLP are hybrid features derived using MFCC and PLP. They used two different models based on Vector Quantization (VQ) with Dynamic Time Warping (DTW) and Gaussian Mixture Model (GMM) for classification. These classification models were trained with different features. The authors were able to show that these models worked better with hybrid features (BFCC and RPLP) as compared to conventional features (MFCC and PLP). GMM combined with RPLP features gave the most promising results and achieved an accuracy of 88.8% on ten languages. They designed their own dataset comprising of ten languages being Dutch, English, French, German, Italian, Russian, Spanish, Hindi, Telugu, and Bengali.

Montavon [17] generated Mel spectrogram as features for a time-delay neural network (TDNN). This network had two-dimensional convolutional layers for feature extraction. An elaborate analysis of how deep architectures outperform their shallow counterparts is presented in this research. The difficulties in classifying perceptually similar languages like German and English were also put forward in this work. It is mentioned that the proposed approach is less robust to new speakers present in the test dataset. This method was able to achieve an accuracy of 91.2% on dataset comprising of 3 languages – English, French and German.

In Table 1, we summarize the quantitative results of the above previous studies. It includes the model basis, feature description, languages classified and the used dataset along with accuracy obtained. The table also lists the overall results of our proposed models (at the top). The languages used by various authors along with their acronyms are English (En), Spanish (Es), French (Fr), German (De), Russian (Ru), Italian (It), Bengali (Ben), Hindi (Hi) and Telugu (Tel).

Table 1. Quantitative review of previous studies along with our results.

Year	Model basis	Features	Languages	Acc.	Remarks	Ref.
2019	1D ConvNet	Raw Audio	En, Fr, De, Es, Ru, It	93.7 ^a	Evaluation of our 1D ConvNet model with mixup for six languages	Self
2019	2D ConvNet	log-Mel	En, Fr, De, Es, Ru, It	95.4 ^a	Evaluation of our 2D ConvNet model with mixup for six languages	Self
2019	2D ConvNet-Bi-directional GRU-Attention	log-Mel	En, Fr, De, Es, Ru, It	95.0 ^a	Result after tuning the hyperparameters of our cnn-bi-directional GRU-attention model and applying mixup	Self
2019	2D ConvNet	log-Mel	En, Fr, De, Es	96.3 ^a	Our evaluation of 2D ConvNet model for four languages	Self
2019	ResNet50	log-Mel	En, Fr, De, Es, Ru, It	89.0 ^a	Uses a pretrained ResNet50 architecture and cyclic learner to identify the language	[19]
2018	SVM-HMM model	not defined	En, Fr, Es, De	70.0 ^a	HMMs were used to encode speech into sequences of vectors which were then fed into a neural network	[8]
2017	Inceptionv3 CRNN	log-Mel	En, Fr, De, Es	96.0 ^b	Used Inception-v3 model followed by bi-directional LSTMs to extract convolutional and temporal features	[1]
2017	CRNN	log-Mel	En, Fr, De, Es	91.0 ^b	A new architecture is used to extract spatial features by using CNNs and temporal features using RNNs	[1]
2010	Gaussian Mixture Models	Perceptual Linear Prediction	Dut, En, Fr, De, It, Ru, Es, Ben, Hi and Tel	88.8 ^c	Used Gaussian mixture models coupled with RPLP features, which were prepared using MFCC and PLP	[12]
2009	CNN-TDNN	log-Mel	En, Fr, De	91.2 ^a	Used a time delay neural network with SGD was used to identify language using log-Mel images as input	[17]

Dataset: ^a - VoxForge [23]; ^b - Youtube News [27], ^c - Private

3 Proposed Method

3.1 Motivations

Several state-of-the-art results on various audio classification tasks have been obtained by using log-Mel spectrograms of raw audio, as features [25]. Convolutional Neural Networks have demonstrated an excellent performance gain in classification of these features [10, 26] against other machine learning techniques. It has been shown that using *attention* layers with ConvNets further enhanced their performance [13]. This motivated us to develop a CNN-based architecture with *attention* since this approach hasn't been applied to the task of language identification before.

Recently, using raw audio waveform as features to neural networks has become a popular approach in audio classification [13, 24]. Raw waveforms have several artifacts which are not effectively captured by various conventional feature extraction techniques like Mel Frequency Cepstral Coefficients (MFCC), Constant Q Transform (CQT), Fast Fourier Transform (FFT), etc.

Audio files are a sequence of spoken words, hence they have temporal features too. A CNN is better at capturing spatial features only and RNNs are better

at capturing temporal features as demonstrated by Bartz et al. [1] using audio files. Therefore, we combined both of these to make a CRNN model.

We propose three types of models to tackle the problem with different approaches, discussed as follows.

3.2 Description of Features

As an average human’s voice is around 300 Hz and according to Nyquist-Shannon sampling theorem all the useful frequencies (0–300 Hz) are preserved with sampling at 8 kHz, therefore, we sampled *raw audio* files from all six languages at 8 kHz.

The average length of audio files in this dataset was about 10.4 s and standard deviation was 2.3 s. For our experiments, the audio length was set to 10 s. If the audio files were shorter than 10 s, then the data was repeated and concatenated. If audio files were longer, then the data was truncated.

3.3 Model Description

We applied the following design principles to all our models:

- *Every convolutional layer is always followed by an appropriate max pooling layer.* This helps in containing the explosion of parameters and keeps the model small and nimble.
- *Convolutional blocks* are defined as an individual block with multiple pairs of one convolutional layer and one max pooling layer. *Each convolutional block is preceded or succeeded by a convolutional layer.*
- *Batch Normalization and Rectified linear unit activations were applied after each convolutional layer.* Batch Normalization helps speed up convergence during training of a neural network.
- Model *ends with a dense layer* which acts the final output layer.

3.4 Model Details: 1D ConvNet

As the sampling rate is 8 kHz and audio length is 10 s, hence the input is *raw audio* to the models with input size of (batch size, 1, 80000). In Table 2, we present a detailed layer-by-layer illustration of the model along with its hyperparameter.

Hyperparameter Optimization: Tuning hyperparameters is a cumbersome process as the hyperparameter space expands exponentially with the number of parameters, therefore efficient exploration is needed for any feasible study. We used the *random search* algorithm supported by *Hyperopt* [2] library to randomly search for an optimal set of hyperparameters from a given parameter space. In Fig. 1, various hyperparameters we considered are plotted against the validation accuracy as violin plots. Our observations for each hyperparameter are summarized below:

Table 2. Architecture of the 1D-ConvNet model

Layer name	# filters/kernel/ stride	output	# of parameters
Conv1	(128, 3, 3)	(128, 26664)	384
(Convolutional Block 1)			
Conv1D	(128, 3, 1)	(128, 26658)	49152
MaxPool1D	(3, 3)	(128, 8880)	
Conv1D	(128, 3, 1)	(128, 8880)	49,152
MaxPool1D	(3, 3)	(128, 2960)	
Conv1D	(256, 3, 1)	(256, 2954)	98,304
MaxPool1D	(3, 3)	(256, 984)	
(Convolutional Block 2)			
Conv1D	(256, 3, 1)	(256, 978)	196,608
MaxPool1D	(3, 3)	(256, 326)	
Conv1D	(512, 3, 1)	(512, 320)	393,216
MaxPool1D	(106, 3)	(512, 1)	
Dense Layer	(512, 6)	(6)	3,072

Number of Filters in First Layer: We observe that having 128 filters gives better results as compared to other filter values of 32 and 64 in the first layer. A higher number of filters in the first layer of network is able to preserve most of the characteristics of input.

Kernel Size: We varied the receptive fields of convolutional layers by choosing the kernel size from among the set of {3, 5, 7, 9}. We observe that a kernel size of 9 gives better accuracy at the cost of increased computation time and larger number of parameters. A large kernel size is able to capture longer patterns in its input due to bigger receptive power which results in an improved accuracy.

Dropout: Dropout randomly turns-off (sets to 0) various individual nodes during training of the network. In a deep CNN it is important that nodes do not develop a co-dependency amongst each other during training in order to prevent overfitting on training data [21]. Dropout rate of 0.1 works well for our model. When using a higher dropout rate the network is not able to capture the patterns in training dataset.

Batch Size: We chose batch sizes from amongst the set {32, 64, 128}. There is more noise while calculating error in a smaller batch size as compared to a larger one. This tends to have a regularizing effect during training of the network and hence gives better results. Thus, batch size of 32 works best for the model.

Layers in Convolutional Block 1 and 2: We varied the number of layers in both the convolutional blocks. If the number of layers is low, then the network does

not have enough depth to capture patterns in the data whereas having large number of layers leads to overfitting on the data. In our network, two layers in the first block and one layer in the second block give optimal results.

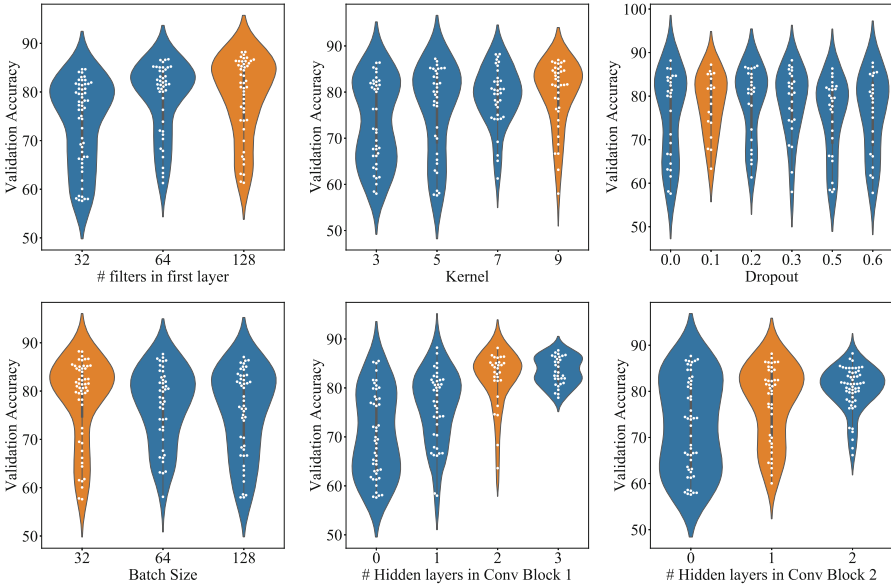


Fig. 1. Effect of hyperparameter variation of the hyperparameter on the classification accuracy for the case of 1D-ConvNet. Orange colored violin plots show the most favored choice of the hyperparameter and blue shows otherwise. One dot represents one sample. (Color figure online)

3.5 Model Details: 2D ConvNet with Attention and Bi-Directional GRU

Log-Mel spectrogram is the most commonly used method for converting audio into the image domain. The audio data was again sampled at 8 kHz. The input to this model was the log-Mel spectra. We generated log-Mel spectrogram using the *LibROSA* [14] library. In Table 3, we present a detailed layer-by-layer illustration of the model along with its hyperparameter.

We took some specific design choices for this model, which are as follows:

- We added *residual connections* with each convolutional layer. Residual connections in a way makes the model selective of the contributing layers, determines the optimal number of layers required for training and solves the problem of vanishing gradients. Residual connections or skip connections skip training of those layers that do not contribute much in the overall outcome of model.

Table 3. Architecture of the 2D-ConvNet model

Layer Name	Output features	Number of filters/stride/padding	No. of parameters
(ConvBlock 1)			
Conv2D	(64, 128, 128)	(3, 3)/(1, 1)/(1, 1)	1,728
Conv2D	(64, 128, 128)	(3, 3)/(1, 1)/(1, 1)	36,864
AvgPool2D	(64, 64, 64)		
(ConvBlock 2)			
Conv2D	(128, 64, 64)	(3, 3)/(1, 1)/(1, 1)	73,728
Conv2D	(128, 64, 64)	(3, 3)/(1, 1)/(1, 1)	147,456
AvgPool2D	(128, 32, 32)		
(ConvBlock 3)			
Conv2D	(256, 32, 32)	(3, 3)/(1, 1)/(1, 1)	294,912
Conv2D	(256, 32, 32)	(3, 3)/(1, 1)/(1, 1)	589,824
AvgPool2D	(256, 16, 16)		
(ConvBlock 4)			
Conv2D	(512, 16, 16)	(3, 3)/(1, 1)/(1, 1)	1,179,648
Conv2D	(512, 16, 16)	(3, 3)/(1, 1)/(1, 1)	235,929
AvgPool2D	(512, 8, 8)		
Bi-directional GRU	(8, 1536)		1,769,472
Embedding Layer (Sequential Block)	(8, 768)		1,179,648
Dropout (0.2) Linear	(256)		131,072
Dropout (0.1) Linear	(6)		1,536

- We added *spatial attention* [3] networks to help the model in focusing on specific regions or areas in an image. Spatial attention aids learning irrespective of transformations, scaling and rotation done on the input images making the model more robust and helping it to achieve better results.
- We added *Channel Attention* networks so as to help the model to find interdependencies among color channels of log-Mel spectra. It adaptively assigns importance to each color channel in a deep convolutional multi-channel network. In our model we apply channel and spatial attention just before feeding the input into bi-directional GRU. This helps the model to focus on selected regions and at the same time find patterns among channels to better determine the language.

Hyperparameter Optimization: We used the *random search* algorithm supported by *Hyperopt* [2] library to randomly search for an optimal set of hyperparameters from a given parameter space. In Fig. 2, various hyperparameters we tuned are plotted against the validation accuracy. Our observations for each hyperparameter are summarized below:

Filter Size: 64 filters in the first layer of network can preserve most of the characteristics of input, but increasing it to 128 is inefficient as overfitting occurs.

Kernel Size: There is a trade-off between kernel size and capturing complex non-linear features. Using a small kernel size will require more layers to capture

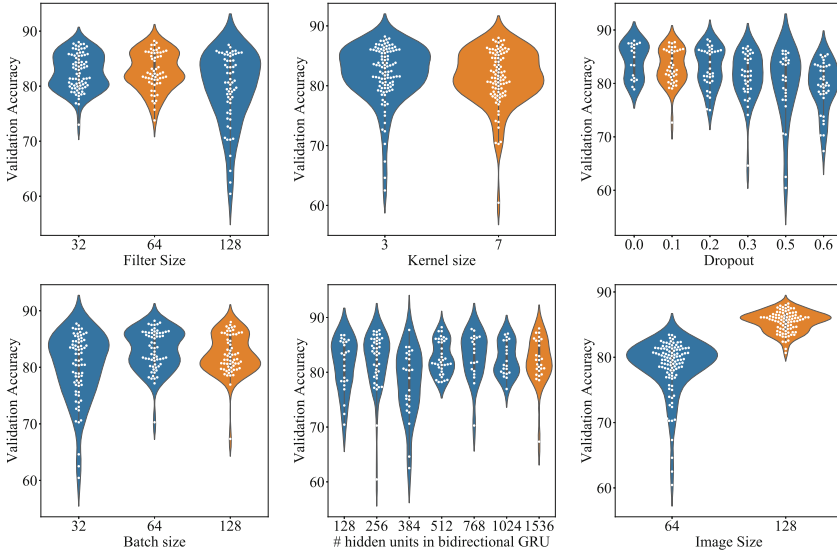


Fig. 2. Effect of hyperparameter variation of the six selected hyperparameter on the classification accuracy for the case of 2D-ConvNet. Orange colored violin plots show the most favored choice of the hyperparameter and blue shows otherwise. One dot represents one sample. (Color figure online)

features whereas using a large kernel size will require less layers. Large kernels capture simple non-linear features whereas using a smaller kernel will help us capture more complex non-linear features. However, with more layers, backpropagation necessitates the need for a large memory. We experimented with large kernel size and gradually increased the layers in order to capture more complex features. The results are not conclusive and thus we chose kernel size of 7 against 3.

Dropout: Dropout rate of 0.1 works well for our data. When using a higher dropout rate the network is not able to capture the patterns in training dataset.

Batch Size: There is always a trade-off between batch size and getting accurate gradients. Using a large batch size helps the model to get more accurate gradients since the model tries to optimize gradients over a large set of images. We found that using a batch size of 128 helped the model to train faster and get better results than using a batch size less than 128.

Number of Hidden Units in Bi-directional GRU: Varying the number of hidden units and layers in GRU helps the model to capture temporal features which can play a significant role in identifying the language correctly. The optimal number of hidden units and layers depends on the complexity of the dataset. Using less number of hidden units may capture less features whereas using large number

of hidden units may be computationally expensive. In our case we found that using 1536 hidden units in a single bi-directional GRU layer leads to the best result.

Image Size: We experimented with log-Mel spectra images of sizes 64×64 and 128×128 pixels and found that our model worked best with images of size of 128×128 pixels.

We also evaluated our model on data with mixup augmentation [28]. It is a data augmentation technique that also acts as a regularization technique and prevents overfitting. Instead of directly taking images from the training dataset as input, mixup takes a linear combination of any two random images and feeds it as input. The following equations were used to prepared a mixed-up dataset:

$$\text{Input_Image} = \alpha * I_1 + (1 - \alpha) * I_2, \quad (1)$$

and

$$\text{Input_Label} = \alpha * L_1 + (1 - \alpha) * L_2, \quad (2)$$

where $\alpha \in [0, 1]$ is a random variable from a β -distribution, I_1 .

3.6 Model Details: 2D-ConvNet

This model is a similar model to 2D-ConvNet with Attention and bi-directional GRU described in Sect. 3.5 except that it lacks skip connections, attention layers, bi-directional GRU and the embedding layer incorporated in the previous model.

3.7 Dataset

We classified six languages (English, French, German, Spanish, Russian and Italian) from the VoxForge [23] dataset. VoxForge is an open-source speech corpus which primarily consists of samples recorded and submitted by users using their own microphone. This results in significant variation of speech quality between samples making it more representative of real world scenarios.

Our dataset consists of 1,500 samples for each of six languages. Out of 1,500 samples for each language, 1,200 were randomly selected as training dataset for that language and rest 300 as validation dataset using k-fold cross-validation. To sum up, we trained our model on 7,200 samples and validated it on 1800 samples comprising six languages. The results are discussed in next section.

4 Results and Discussion

This paper discusses two end-to-end approaches which achieve state-of-the-art results in both the image as well as audio domain on the VoxForge dataset [23]. In Table 4, we present all the classification accuracies of the two models of the cases with and without mixup for six and four languages.

In the audio domain (using raw audio waveform as input), 1D-ConvNet achieved a mean accuracy of 93.7% with a standard deviation of 0.3% on running k-fold cross validation. In Fig. 3(a) we present the confusion matrix for the 1D-ConvNet model.

In the image domain (obtained by taking log-Mel spectra of raw audio), 2D-ConvNet with 2D attention (channel and spatial attention) and bi-directional GRU achieved a mean accuracy of 95.0% with a standard deviation of 1.2% for six languages. This model performed better when mixup regularization was applied. 2D-ConvNet achieved a mean accuracy of 95.4% with standard deviation of 0.6% on running k-fold cross validation for six languages when mixup was applied. In Fig. 3(b) we present the confusion matrix for the 2D-ConvNet model. 2D attention models focused on the important features extracted by convolutional layers and bi-directional GRU captured the temporal features.

Table 4. Results of the two models and all its variations

Languages	Feature desc.	Network	Mixup	Accuracy
En, Es, Fr, De, Ru, It	Raw Waveform	1D ConvNet	No	93.7
	log-Mel Spectra	2D ConvNet	No	94.3
			Yes	95.4
		2D ConvNet with Attention and GRU	No	94.3
			Yes	95.0
	En, Es, Fr, De	Raw Waveform	1D ConvNet	No
log-Mel Spectra		2D ConvNet	No	96.0
			Yes	96.3
		2D ConvNet with Attention and GRU	No	94.7
			Yes	93.7

4.1 Misclassification

Several of the spoken languages in Europe belong to the Indo-European family. Within this family, the languages are divided into three phyla which are Romance, Germanic and Slavic. Of the 6 languages that we selected Spanish (Es), French (Fr) and Italian (It) belong to the Romance phyla, English and German belong to Germanic phyla and Russian in Slavic phyla. Our model also confuses between languages belonging to the similar phyla which acts as an insanity check since languages in same phyla have many similar pronounced words such as *cat* in English becomes *Katze* in German and *Ciao* in Italian becomes *Chao* in Spanish.

Our model confuses between French (Fr) and Russian (Ru) while these languages belong to different phyla, many words from French were adopted into Russian such as *automate* (oot-oo-mate) in French becomes *ABTOMaT* (aff-taa-maat) in Russian which have similar pronunciation.

Future Scope. The performance of raw audio waveforms as input features to ConvNet can be further improved by applying silence removal in the audio. Also,

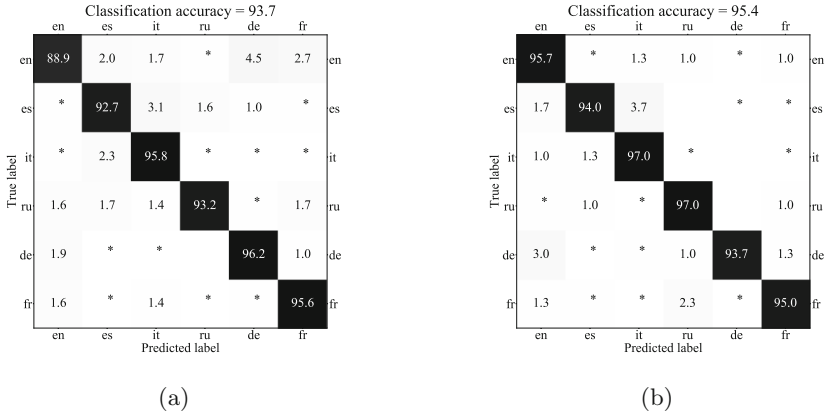


Fig. 3. Confusion matrix for classification of six languages with our (a) 1D-ConvNet and (b) 2D-ConvNet model. Asterisk (*) marks a value less than 0.1%.

there is scope for improvement by augmenting available data through various conventional techniques like pitch shifting, adding random noise and changing speed of audio. These help in making neural networks more robust to variations which might be present in real world scenarios. There can be further exploration of various feature extraction techniques like Constant-Q transform and Fast Fourier Transform and assessment of their impact on Language Identification.

There can be further improvements in neural network architectures like concatenating the high level features obtained from 1D-ConvNet and 2D-ConvNet, before performing classification. There can be experiments using deeper networks with skip connections and Inception modules. These are known to have positively impacted the performance of Convolutional Neural Networks.

5 Conclusion

There are two main contributions of this paper in the domain of spoken language identification. Firstly, we presented an extensive analysis of raw audio waveforms as input features to 1D-ConvNet. We experimented with various hyperparameters in our 1D-ConvNet and evaluated their effect on validation accuracy. This method is able to bypass the computational overhead of conventional approaches which depend on generation of spectrograms as a necessary pre-processing step. We were able to achieve an accuracy of **93.7%** using this technique.

Next, we discussed the enhancement in performance of 2D-ConvNet using mixup augmentation, which is a recently developed technique to prevent overfitting on test data. This approach achieved an accuracy of **95.4%**. We also analysed how *attention* mechanism and recurrent layers impact the performance of networks. This approach achieved an accuracy of **95.0%**.

References

1. Bartz, C., Herold, T., Yang, H., Meinel, C.: Language identification using deep convolutional recurrent neural networks. In: Liu, D., Xie, S., Li, Y., Zhao, D., El-Alfy, E.S. (eds.) *Neural Information Processing*. LNCS, vol. 10639, pp. 880–889. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70136-3_93
2. Bergstra, J., Yamins, D., Cox, D.D.: Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures (2013)
3. Chen, L., et al.: Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5659–5667 (2017)
4. Dehak, N., Torres-Carrasquillo, P.A., Reynolds, D., Dehak, R.: Language recognition via i-vectors and dimensionality reduction. In: *Twelfth Annual Conference of the International Speech Communication Association* (2011)
5. Endah Safitri, N., Zahra, A., Adriani, M.: Spoken language identification with phonotactics methods on minangkabau, sundanese, and javanese languages. *Procedia Comput. Sci.* **81**, 182–187 (2016). <https://doi.org/10.1016/j.procs.2016.04.047>
6. Ferrer, L., Scheffer, N., Shriberg, E.: A comparison of approaches for modeling prosodic features in speaker recognition. In: *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4414–4417. IEEE (2010)
7. Ganapathy, S., Han, K., Thomas, S., Omar, M., Segbroeck, M.V., Narayanan, S.S.: Robust language identification using convolutional neural network features. In: *Fifteenth Annual Conference of the International Speech Communication Association* (2014)
8. Gazeau, V., Varol, C.: Automatic spoken language recognition with neural networks. *Int. J. Inf. Technol. Comput. Sci. (IJITCS)* **10**(8), 11–17 (2018)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016*, pp. 770–778 (2016). <https://doi.org/10.1109/CVPR.2016.90>
10. Hershey, S., et al.: CNN architectures for large-scale audio classification. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 131–135. IEEE (2017)
11. Howard, J., et al.: Fastai (2018). <https://github.com/fastai/fastai>
12. Kumar, P., Biswas, A., Mishra, A.N., Chandra, M.: Spoken language identification using hybrid feature extraction methods. arXiv preprint [arXiv:1003.5623](https://arxiv.org/abs/1003.5623) (2010)
13. Lee, J., Kim, T., Park, J., Nam, J.: Raw waveform-based audio classification using sample-level CNN architectures. arXiv preprint [arXiv:1712.00866](https://arxiv.org/abs/1712.00866) (2017)
14. LibROSA: <https://librosa.github.io/librosa/>. Accessed 16 July 2019
15. Lopez-Moreno, I., Gonzalez-Dominguez, J., Plchot, O., Martinez, D., Gonzalez-Rodriguez, J., Moreno, P.: Automatic language identification using deep neural networks. In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5337–5341. IEEE (2014)
16. Martinez, D., Plchot, O., Burget, L., Glembek, O., Matějka, P.: Language recognition in i-vectors space. In: *Twelfth Annual Conference of the International Speech Communication Association* (2011)
17. Montavon, G.: Deep learning for spoken language identification. In: *NIPS Workshop on Deep Learning for Speech Recognition and Related Applications*, pp. 1–4 (2009)

18. Obuchi, Y., Sato, N.: Language identification using phonetic and prosodic HMMs with feature normalization. In: Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2005), vol. 1, pp. 1–569. IEEE (2005)
19. Revay, S., Teschke, M.: Multiclass language identification using deep learning on spectral images of audio signals. arXiv preprint [arXiv:1905.04348](https://arxiv.org/abs/1905.04348) (2019)
20. Tong, R., Ma, B., Zhu, D., Li, H., Chng, E.S.: Integrating acoustic, prosodic and phonotactic features for spoken language identification. In: 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, vol. 1, p. I, May 2006. <https://doi.org/10.1109/ICASSP.2006.1659993>
21. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014)
22. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)
23. [voxforge.org](http://www.voxforge.org/): Free speech recognition (Linux, Windows and mac) - <http://www.voxforge.org/>. Accessed 16 July 2019
24. Wei, Q., Liu, Y., Ruan, X.: A report on audio tagging with deeper CNN, 1D-convnet and 2D-convnet
25. Xu, K., et al.: General audio tagging with ensembling convolutional neural networks and statistical features. *J. Acoust. Soc. Am.* **145**(6), EL52–EL527 (2019)
26. Xu, Y., et al.: Unsupervised feature learning based on deep models for environmental audio tagging. *IEEE/ACM Trans. Audio Speech Lang. Process.* **25**(6), 1230–1241 (2017)
27. Youtube: <http://www.youtube.com>. Accessed 16 July 2019
28. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: beyond empirical risk minimization. arXiv preprint [arXiv:1710.09412](https://arxiv.org/abs/1710.09412) (2017)
29. Zissman, M.A.: Comparison of four approaches to automatic language identification of telephone speech. *IEEE Trans. Speech Audio Process.* **4**(1), 31 (1996)



Indoor Air Quality and Wellbeing - Enabling Awareness and Sensitivity with Ambient IoT Displays

Andreas Seiderer^(✉), Ilhan Aslan, Chi Tai Dang, and Elisabeth André

Augsburg University, Universitätsstr. 6a, 86159 Augsburg, Germany
{seiderer, aslan}@hcm-lab.de, {dang, andre}@informatik.uni-augsburg.de

Abstract. The quality of indoor air exerts influence on the wellbeing of people. However, people rarely notice a constant and creeping deterioration of indoor air. Especially in enclosed places where several people get together, like meeting rooms, school rooms and public transportation, bad air quality might cause a reduction of cognitive performance, increased headache, fatigue and sleepiness. This paper describes and discusses a privacy respecting system, built with low cost IoT components and open-source software, that informs room occupants about bad air quality with ambient lights. Additionally, the status of the windows is indicated so that they are not forgotten to be closed after airing. We conducted a workshop with an implementation of the system with five users and present the results which show the usefulness and desirability of ambient notifications for indoor air quality monitoring.

Keywords: Ambient display · Ambient light · Air quality · Smart home · Context-aware system · Human wellbeing · Ubiquitous computing

1 Introduction

The indoor room climate plays an important role for human wellbeing, especially as humans do not directly and consciously perceive gases such as carbon dioxide (CO₂). It is therefore worrying that higher concentrations of CO₂ can reduce perceived air quality, potentially affect cognitive performance [11, 13, 33, 36], or cause physiological symptoms [13, 26]. Even perceivable gases are problematic since room occupants usually start to adapt to odors that slowly occur [20]. Such odors usually belong to volatile organic compounds (VOCs) which include gases emitted by humans (bioeffluents) but can for example also stem from cleaning agents. When entering a room, such odors are usually noticed and perceived as disturbing, whereby certain VOCs can also be harmful at higher concentrations.

Current sensor technologies can detect CO₂ and VOCs, allowing to detect when a room should be ventilated in order to create a healthier environment. Buildings that have ventilation/air conditioning systems could automatically

control the airflow in the building to reduce such gases. However, this is not standard in buildings and cannot simply be retrofitted. It therefore makes sense to notify people in a room about poor air quality in a non-disturbing way. Unobtrusive notifications can be achieved with ambient displays at the periphery of a room that are visible to all persons.

For this purpose, we developed a prototype that makes use of ambient lights to inform people in a meeting room about air quality and open windows. The ambient lights, located near each window handle and next to the door, encode air quality with the well-known colors of traffic lights. Furthermore, a slowly pulsed light pattern indicates that a window is still open. The pulse pattern occurs on the light at the door together with the lights at the open window(s) to prevent forgetting open windows when leaving the room. To make the lights less obtrusive the maximum brightness is adapted to the current lighting condition in the room. In addition, the lights are switched on only if there are room occupants in order to save energy.

Our prototype features privacy, flexibility, and reproducibility at low costs. The costs of the presented setup are ca. 440 excluding the WiFi router (often already available), the SCD30, BMP280 and iAM sensors and the Arduino (just required for sensor comparison). High level of privacy is achieved by the sensors chosen and by processing and storing data only locally. It is not possible and not intended to determine data of an individual and a connection to the internet or a cloud is not required even though research has shown that cloud usage in such contexts might be accepted [16]. The system's hardware components are inexpensive, makes use of open-source software, and can be installed in any room by means of wireless smart home sensors and custom-built WiFi controlled RGB-LED actuators in addition to gas sensors connected to a Raspberry Pi.

In this paper we sum up related work and the resulting design decisions for our prototype and describe its implementation and installation in a meeting room of a lab at a university. We present insights on the data of multiple sensors that can be measured in this room. Furthermore, we report and discuss the results of a workshop conducted with five staff members that are using the room regularly. The descriptions and results of this paper can be used to implement comparable low-cost systems that could be installed in every room enabling room occupants to detect imperceptible gases and prevent a reduction of their wellbeing.

2 Related Work

2.1 Influence of CO₂ and Bioeffluents on Wellbeing and Cognitive Performance

In order to determine color encodings for different air quality measures we reviewed literature for reasonable thresholds. The room in which our system is installed in is used for meetings where it is usually required that all persons are highly concentrated. Thus, we are especially interested in literature giving hints about thresholds where already slight reductions of the cognitive performance or wellbeing could be found.

We focus the discussion on CO₂ and neglect VOC concentrations for the following reason. The low-cost VOCs sensor that we included has to be manually calibrated against a reference of “good” air quality (well aired room). In turn, the absolute readings of the sensor cannot be reliably used for comparison with values found in literature as the same value may refer to different qualities depending on the calibration. Nevertheless, this sensor still can provide helpful hints about odors in the room for data analysis.

A recent literature survey by Azuma et al. [13] gives an overview of work related to the effects of CO₂ on the human body. In the following we use this survey as a base for our selection and analysis of papers being relevant for our prototype. A reduction of cognitive performance scores could be measured in a controlled study by Satish et al. [33] with a CO₂ concentration of 1000 ppm after 2.5 h. A reduction was observed by Allen et al. [12] in their controlled study with a CO₂ concentration of about 950 ppm. Maula et al. [26] reported a slightly reduced cognitive performance of their controlled study participants at a CO₂ concentration of 2260 ppm after four hours. The perceived workload and fatigue was increased. However, they could not find a strong effect of human bioeffluents on the work performance.

Zhang et al. [36] could not observe a significant reduction of cognitive performance with increased CO₂ concentrations of up to 3000 ppm during their 4.25 h long controlled experiments. Nevertheless, if bioeffluents were also present at CO₂ concentrations of up to 3000 ppm the cognitive performance was reduced. Additionally, their study participants reported the air quality as being lower and that they experienced increased intensity of headache, fatigue, sleepiness and difficulty in thinking. Zhang et al. supposed that they might not have been able to measure a dependency between cognitive performance and CO₂ concentration due to their simpler cognitive tests in comparison to the strategic management simulation test that was for example used by Allen et al. [12] and Satish et al. [33].

In a more recent study by Allen et al. [11] investigated the performance of commercial airline pilots during a simulated three hours long flight with varied CO₂ concentrations. Especially at a concentration of 2500 ppm the performance of the pilots was significantly lower compared to 700 ppm. At 1500 ppm a minor performance reduction could be observed.

Ratings of CO₂ concentrations can also be found in official documents released by governments. For example in a proclamation of the German “Umweltbundesamt” [10], which is based on several studies before 2008, a CO₂ concentration below 1000 ppm is considered as inconspicuous, between 1000 ppm and 2000 ppm as conspicuous and above 2000 ppm as unacceptable. Comparable guidelines are released by other countries’ governmental institutions.

The results of our literature review shows no exact thresholds for our system. Nevertheless, the recommendations by the German “Umweltbundesamt” from 2008 are still in line with more recent literature. Therefore, we use these thresholds for our system: 1000 ppm as bad and 2000 ppm as very bad air quality.

2.2 Systems Monitoring Indoor Air Quality

Marques et al. presented in [22] the “iAirCO₂” system which monitors indoor air quality. It uses an ESP8266 microcontroller with WiFi connectivity interfacing a NDIR-based (Nondispersive infrared sensor) CO₂ sensor. The data can be accessed via an iOS smartphone app and a web interface. It can inform the user about bad air quality via e-mail, SMS or smartphone notifications. With our developed prototype we use also NDIR-based CO₂ sensors and ESP8266 controllers are used to control the LED stripes wirelessly. For the system output we focus on unobtrusive notifications provided by ambient light that is observable by all persons inside the room. In our case there is no automatic ventilation available and no (external) person is responsible to sufficiently air the room so that the room occupants should be enabled by our system to air the room by themselves. Additionally, we try to prevent open windows to be forgotten when everybody leaves the room.

In the article [30] by Perez et al. an indoor environmental quality monitoring system for schools is being described which has been evaluated in five rooms. The CO₂ concentration determined by a novel NDIR-based sensor, humidity and temperature of a sensor node were transmitted to a gateway via the Z-Wave wireless protocol. The observed rooms didn't provide automated heating, ventilation or air conditioning. The system was used to collect and analyze data and provides estimations how long a room should be aired. The authors included the humidity and temperature as these also affect human wellbeing. They conclude that it is necessary to use air monitors to be able to sufficiently air rooms at a school. Their system could be a low-cost solution for this problem. In this study the author's system did not directly notify the students or teachers or let them access the data. Thus, particularly the notification component of our developed system could be an addition to their sensor system. In comparison to Perez et al. we didn't take temperature and humidity into account as they are more likely to be noticed and adapted by airing or controlling the heating. Nevertheless, for a future prototype this might be a possible addition where also energy efficient airing could be considered.

A system intended to prevent sudden infant death which additionally to other data observes the indoor air quality is presented by De La Iglesia et al. in [18]. To observe several gas concentrations multiple low-cost gas sensors based on the MOS (Metal Oxide Semiconductor) principle are used which are interfaced by an ESP8266 microcontroller. In comparison to NDIR-based sensors the readings of MOS sensors usually include the gas concentrations of other gases which should be considered when interpreting the data. For this reason we use NDIR-based CO₂ sensors for our system which are less influenced by other gases. To measure VOCs a combined value of several gases is intended and thus for this purpose MOS-based sensors are usually the better choice. For notifications rules are used that determine different severity levels of a problematic situation depended on the context. A mobile app is included to notify parents and caregivers. For emergency situations ambient lights are not intended to be used as they are more likely to be overlooked, nevertheless they can usually easily be switched

to act obtrusively. In our case increased brightness or eye-catching brightness patterns like fast flashing would be possible. Nevertheless, the meaning of the encoding would have to be made clear.

A commercial device to monitor the air quality at home can be found with the “idevices” “Air Quality Sensor” [1]. It measures VOCs, CO₂, temperature, humidity and air pressure and is connected to the company’s cloud service where it stores the data. The product most likely incorporates the Bosch BME680 which is MOS-based using the compact MEMS fabrication (Micro-Electro-Mechanical System). The cloud service of “idevices” provides a web interface and smartphone app to access the sensor data. It can be integrated in several smart home systems. The sensor itself incorporates LEDs to indicate very good, alarming, and bad air quality. This commonly used traffic light color encoding is also applied to our system. The device itself does not provide hints what purpose it has so that the color of the LEDs is not directly understandable. We try to reduce this mapping problem by locating the ambient lights next to the windows so that there is a tight association with airing and windows status. As we use several LED stripes at least one of them is usually in the field of sight of persons inside the room so that it is very likely that at least one occupant notices the ambient notifications. All data of our system is stored inside the same room so that a high level of control and privacy can be guaranteed.

2.3 Guidelines and Categorization of Ambient Light Systems

For ambient information systems several guidelines and classifications can be found. Nevertheless, there are just few for ambient light systems which are explicitly dedicated for the modality “light”.

In the work [25] by Matviienko et al. nine guidelines derived from their experimental results are listed. For our system we used the traffic light color encoding due to their guideline “GL4”. This color mapping was found as being most frequently proposed by their study participants for everyday situations. According to traffic lights and “GL1” red is a negative and green a positive indicator.

Matviienko et al. additionally proposed a categorization of ambient lights in [24] and applied it to 72 ambient light systems. Their categorization is derived from several guidelines and heuristics for ambient information systems that can be found in the literature.

In the following we classify our system according to their work. Terms used by Matviienko et al. to classify systems are written in italics. Our system is showing the current *status*. *Notifications* which could distract the users are not implemented in our system since the provided information is never critical. Thus, we use color changes and just a slowly pulsed light pattern. The information of our system is encoded by the *color* (air quality), *brightness* pattern (window state) and *LED position* (right beside window handles/door). In the *office* is the *context* where our system is being used. The *information sensitivity* is *public* since several persons inside the room can see the LEDs. From the technical perspective we use *multiple LEDs* which can provide *multicolor* light. Our system is *stationary*.

2.4 Ambient Light in the Office Environment

Ambient light can be used in several contexts. In this case we focus on the office context. Nevertheless, due to the usage of our system in a meeting room there are some identical requirements to be met which occur in the home context. For example, it is required to dim down ambient lights in low light conditions so that they don't become disturbing. In the home environment this can occur in the evening and night while the inhabitants are watching TV if an ambient light is placed nearby. This problem for example had to be dealt with in the work by Seiderer et al. [34] and could be observed for an ambient screen by Consolvo and Towle in [14]. The same issue occurs in the meeting room during presentations using the projector in the evening or when the room has to be darkened with the indoor and/or outdoor shutters. This is why we include an adaptation of the LEDs' brightness according to the brightness measured in the room.

Health related ambient light systems, trying to increase the amount of movement of individual office workers, were for example presented by Fortmann et al. [19] and Mateevitsi et al. [23]. The light sources were installed nearby the workplaces. To provide biofeedback with ambient light has also been researched in the work by Yu et al. [35] where it is used for stress intervention and relaxation. Two works by Müller et al. use ambient lights at the back of the PC monitor of workers. In the first work it is applied to remind them about upcoming events and appointments [28], in the second to let them keep track of secondary tasks in general [27]. The mentioned systems for individuals could be an addition to our system and can also improve the wellbeing of the workers or their time planning.

An ambient light system used in a meeting room to support the time management has been researched by Occhialini et al. [29]. Additionally, other systems can be found that are intended to provide unobtrusive information during presentations so that the course of a meeting can be adapted accordingly. Our system should not intervene with the meetings and just provide feedback so that the room can be aired adequately by the persons preventing a reduction of wellbeing and cognitive performance of the participants due to bad air quality.

3 Implementation

3.1 Hardware

Figure 1 shows an overview of the hardware that we employed. A Raspberry Pi 3 B+ connected via WiFi served as the central device that runs all the software that is necessary to acquire, process and store sensor data and to control LED actuators based on rules. Our prototype included two different low budget CO₂ sensors and a VOCs sensor in order to compare their readings, their dependency on temperature, humidity, and air pressure and observe general sensor drifts. A final setup requires just one reliable sensor.

Figure 1 further shows that an Arduino Mini Pro is connected to the Raspberry. It is supplied with 3.3V at 8MHz and interfaces via I²C a Sensirion

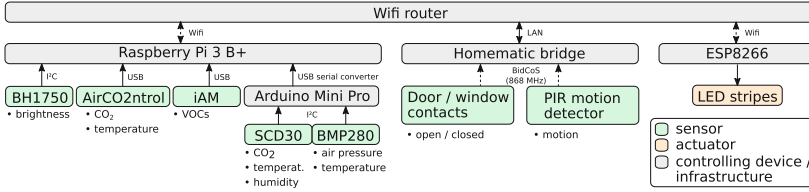


Fig. 1. Hardware overview of the prototype.

SCD30 [2] NDIR based calibrated CO₂ sensor and a Bosch BMP280 [3] air pressure sensor. The SCD30 incorporates a temperature and humidity sensor which allows the sensor’s firmware to compensate the dependencies on temperature and humidity. An additional compensation on air pressure is possible for which the BMP280 sensor would be used. The Arduino, SCD30 and BMP280 components are enclosed in a 3d printed case with airing slots (Fig. 2).

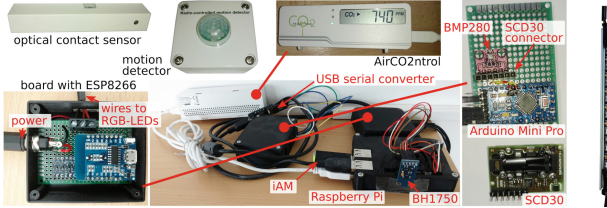


Fig. 2. The picture shows the Raspberry Pi 3 B+ with the gas, temperature, humidity, air pressure, light sensors located next to the door on a sideboard. Additionally, one of the three identical ESP8266 boards for the RGB-LED stripes, the Homematic motion detector and one of the six Homematic optical contact sensor are visible. On the right one of the six LED-stripes with its eight LEDs glued to an aluminum profile is shown without the translucent shade.

A TFA Dostmann AirCO2ntrol Mini [4] CO₂ sensor, which includes a temperature reading, is connected to the Raspberry Pi via USB. This sensor is built as a consumer device and more cost-effective (~ 40 €) than the more recent SCD30 (~ 60 €). We placed the sensor so that the display and the LEDs are not visible to room occupants. According to data sheets, the SCD30 is able to measure CO₂ concentrations in a range of 400 to 10,000 ppm (calibrated) while the AirCO2ntrol sensor is capable of a range of 0 to 3000 ppm.

Furthermore, an ams iAM [5] VOCs sensor is connected via USB to the Raspberry Pi. As typical for VOCs sensors, it delivers values in ppm which are relative to the air quality at power-up or a previously stored value.

To get information about the status of windows and door, and movement in the room, we employ battery powered sensors of the Homematic [6] wireless system. At each window and door, an optical contact sensor (HM-SEC-SCo) is placed. One motion sensor (HM-SEN-MDIR-SM) is located at the side of

the room. The sensor data can be received by means of a bridge device that is connected to the router via LAN.

The ambient brightness in lux in the room is determined by a BH1750 brightness sensor connected to the Raspberry Pi via I²C.

As actuators, we are using six (door: 1, windows: 5) RGB-LED stripes consisting out of eight WS2812B LEDs (60 LEDs per meter) each which are controlled via three ESP8266 microcontrollers. Each stripe is encased in an aluminum profile with a translucent plastic shade. Since high currents are required if the LEDs are running with white color and full brightness (max. 60 mA per LED), 5 V power supplies with 2.5 A maximum current are used to prevent unexpected behavior of the controllers or LEDs due to power breakdowns. A maximum of three stripes (24 LEDs) is connected to one ESP8266 and power supply. The microcontrollers are connected to the WiFi. The ESP8266 board with connectors and a level shifter (5 V ↔ 3.3 V) are protected by a 3d printed case.

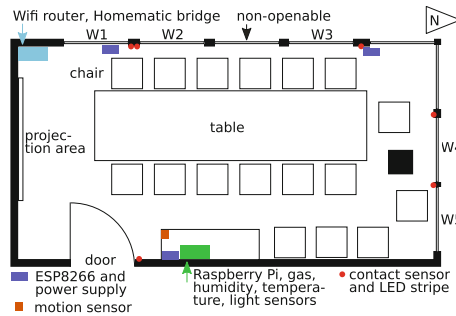


Fig. 3. Overview of the meeting room with hardware positions. Openable windows are labeled with W_x where x is a number. Positions and lengths in general and the count of chairs/non-openable windows are not exact.

3.2 Software

An overview of the communication between the software is visible in Fig. 4.

On the **Raspberry Pi** Raspbian Linux is installed. For communication between some software components the Mosquitto MQTT broker is used. For data storage and partly also processing PostgreSQL is set up. The lightweight open-source home automation software FHEM [7] interfaces the Homematic sensors, reads out the sensor data of the iAM USB stick, acquires the brightness data of the BH1750 and receives all other sensor data via MQTT. Additionally, it is connected with the software on the ESP8266 controllers to be able to control them. We adapted a code to read out the sensor data of the AirCO2ntrol device so that it can transmit data via MQTT. It is shared on Github¹. Grafana is used to visualize and explore data in the database in real time.

¹ <https://github.com/andreas-seiderer/co2mon>.

To receive data from the Arduino Mini Pro (for event processing and context interpretation), we use our own open-source software called Eventerpretor² which was already used in [32] and has been extended since then. Our software is written in Kotlin (instead of native or hybrid frameworks, e.g., [15]), can be used on Windows and Linux, and improves and extends the functionality of the single threaded FHEM for this project. It additionally allows replacing FHEM or including extensions without the need of major changes of the rules. This is especially beneficial for prototyping. A list of benefits in comparison to FHEM relevant for this project is shown in the following:

- flow-based programming with several reusable nodes,
- multi-threaded processing (the Raspberry Pi’s 4 CPU cores can be utilized),
- database provided data aggregations,
- independent sensor model and rule engine (Drools),
- caching/synchronization of the rule engine’s facts,
- custom web interfaces and replay function of recorded events for testing,
- generic serial port interfacing.

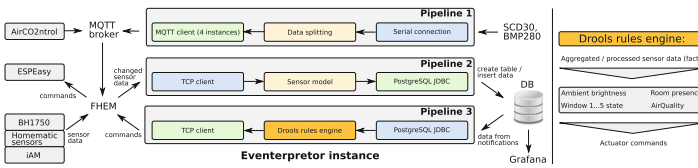


Fig. 4. Software processing and communication overview.

Some functionality is comparable to the more complex Node-RED by IBM for Node.js, nevertheless Eventerpretor includes for our purposes required missing features. Due to the use of Java, we can use real multi-threading for each node and integrate any Java library.

Three pipelines are running in one instance of Eventerpretor. They are visible in Fig. 4. The **first pipeline** receives data of the Arduino Mini Pro from the serial port and processes them (e.g. value separation) so that they can be sent via MQTT to FHEM.

The **second pipeline** directly receives all sensor events of FHEM. FHEM is configured in a way that only transitions of different values are processed as new events. The events of FHEM are parsed by Eventerpretor and supplemented with metadata such as sensor type or sensor position. This additional data is used to automatically create tables with sensor data in the database and to add new sensor entries in a “sensors” table where the additional sensor information can be accessed and the related tables are referenced. Since Eventerpretor is implementing its own sensor model, it is possible to use any other program that can interface Homematic devices and the iAM sensor.

² <https://github.com/andreas-seiderer/Eventerpretor>.

The **third pipeline** receives processed sensor data from the database. This is implemented in a way that it creates trigger functions that can aggregate sensor data over time. For example, the average sensor value over the last five minutes can be calculated whenever a new sensor value has been inserted by the second pipeline. After the calculation, a notification is sent to all subscribers including the calculated value - in this case Eventpreter listens to the trigger that it has created. The possibility of asynchronous messaging is a special function of PostgreSQL and can be very useful to prevent inefficient polling loops. After receiving the aggregated values, the pipeline uses the data with the sensor model (facts) with the rule engine “Drools” [8]. The rule system first generates context elements like the interpretation of the air quality out of the CO₂ concentration and finally creates the commands (cf. right side of Fig. 4). The commands are fed to FHEM which controls the different RGB-LED stripes. Just new commands are sent to prevent a high CPU load of the ESP8266 controllers.

Eventpreter can directly host websites that allow simulating sensor data via a websocket connection but also can be used to receive and visualize events. This is useful during the development and prototyping process. Additionally, received events can be recorded and replayed to simulate specific situations like entering the room and opening a window with realistic sensor data and timing.

The air pressure is acquired by the **Arduino Mini Pro** firmware from the BMP280 sensor and directly sent to the SCD30 firmware for data compensation. After receiving the data from the SCD30 the Arduino sends all its readouts (CO₂ in ppm, temperature in °C, humidity in %) and BMP280 (air pressure in hPa and temperature in °C) to the Raspberry Pi via a USB serial converter.

The **ESP8266** boards are flashed with ESPEasy [9]. This firmware allows an easy setup of the controller via a webinterface. It can be used to connect it to a password protected WiFi without having to hard code the SSID and password. Additionally, it can interface sensors but also actuators. In our case we set up the WS2812B LED stripes which then can be controlled over the network. Since we required new brightness patterns we created a new plugin for ESPEasy.

3.3 Setup

In Fig. 3 the positions of the hardware are visible. In Fig. 5 pictures of LED-stripes next to the door (a) and four of the five windows (b) and (c) are visible. The room has a volume of about 93.4 m³ and is used by less than 20 persons at the same time. Through W1, W2 and W3 and non-openable windows it is possible that direct sunlight can enter the room. W3 and W4 are directed to the north.

If the projector is used and there is too much sunlight shining into the room, indoor and electrical outdoor shutters can be closed. The indoor shutters can manually be moved with a cord horizontally and the vertical angle of the shutter’s stripes can be adapted, too. These shutters are located in front of the LED stripes so that they are no longer in sight if they are completely closed. In this case just the LED stripe at the door is directly visible.

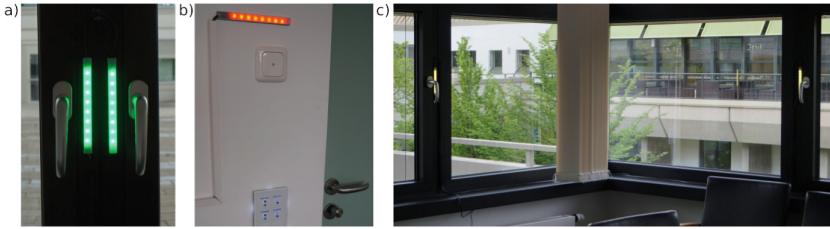


Fig. 5. (a) Shows the LED-strips next to window W1 and W2. In (b) the stripe next to the door can be seen. In (c) the stripes next to window W3 and W4 are visible. Additionally, two non-openable windows and one of the manual indoor shutters can be noticed. For demonstration purposes all three colors are shown in the pictures.

The Raspberry Pi including the gas sensors is placed on a sideboard which is about 110 cm high so that the CO_2 concentration is measured at about the usual height of sitting persons. The room has five windows that can be opened and one door which leads into a corridor. At the side of each window handle a LED stripe is placed. At each window and door a contact sensor is added.

4 Evaluation

For a basic evaluation of ambient displays the heuristics by Mankoff et al. [21] can be used to identify major problems already during the development process. Three instructed persons applied the heuristics on our system. As a result we changed the pulsed brightness pattern so that it doesn't completely dim down to dark but to a lower brightness since it was considered as being too obtrusive.

As a first step, we used the sensors to record data without providing any notifications. This phase of the system was used to test and improve the stability of the prototype. Additionally, the recorded data shows what gas concentrations are to be expected in this room, how often it is being aired and how the different gas sensors behave. After that we activated the notifications and conducted a workshop to gather further insights about the system.

4.1 Data Recording

First data recordings were conducted during winter and spring 2018/2019. The system is intended to gather long-term data to allow more detailed analysis in the future. Figure 6 shows example data which we discuss in more detail.

The visible data of the meeting room was recorded in February where a maximum outside temperature of -1.6°C was reached. At the top the gas concentrations of CO_2 and VOCs are visible provided by the three gas sensors. Below the temperature readings of the AirCO2ntrol sensor are shown. After that the state of the door and the five windows (W1–W5) is plotted.

The CO_2 readings of the AirCO2ntrol and SCD30 sensors just slightly differ. The AirCO2ntrol data shows in general a more noisy signal. At the selected

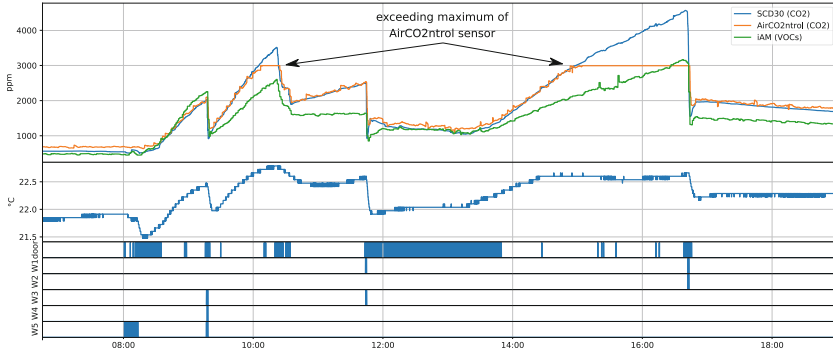


Fig. 6. Example gas sensor measurements in the meeting room of a day in February 2019. The VOCs (related to “fresh” air) and CO₂ (factory calibration) concentrations can just be compared by their relative value changes. In this case no higher concentrations of gases next to CO₂ could be detected by the iAM sensor (this would have resulted in peaks only shown by this sensor). The maximum outside temperature was around -1.6°C . Drop downs in CO₂ and VOCs concentrations indicate an open door or open windows (W1–W5). The CO₂ measurement range of the AirCO2ntrol sensor was exceeded two times.

day the maximum range of the sensor has been exceeded two times so that for this room it is necessary to use sensors with a higher maximum range like the SCD30 sensor to be able to observe the maximum CO₂ concentrations. The VOCs concentration is strongly depended on the CO₂ concentration and in this case shows similar curve shapes. Nevertheless, this is not always true and in such cases indicates the presence of other gases. Although the VOCs sensor is just roughly calibrated it could be useful for our system to be able to notify about other potentially unpleasant gases in the room air.

The temperature measured by the AirCO2ntrol sensor didn’t change much during the day (in a range of about 1.5°C). The gas concentration values and the temperature responded quite quickly to open windows. As soon as just the door was open, the speed of the gas concentration drop is lower.

In general, it can be seen that the room was aired multiple times but not often and partly not long enough so that relatively high gas concentrations could be measured inside the room. One reason for too short airing might have been the lower temperatures outside but more likely the air was no longer noticed as being bad. Especially during the long person presence around 14 o’clock the room occupants did not recognize the bad air quality and the room should have been aired. Although the door was partly opened for longer time periods, this just slightly reduced the VOCs and CO₂ concentrations or caused them to stay at a relatively high level.

During the total recorded time period in winter and spring there are several days when bad and very bad air quality was reached inside the room but very rarely with gas concentrations above 3000 ppm. Thus, with activated ambient display our system has high potential in improving this situation.

4.2 Workshop with Meeting Room Users

In order to gain insights from staff members who have been using the meeting room on a regular basis we conducted a workshop. Workshop participants were five staff members with expertise in human-robot interaction, machine learning, or mobile applications. Three researchers conducted the workshop. One researcher facilitated the workshop and two additional researchers took notes. The workshop took place in the meeting room where the system was installed. Initially, the five participants were given a brief description to the system in room, including a reminder/demonstration of how the system’s ambient feedback changes when the air quality gets worse and when it gets better.

After that, the facilitator guided a semi-structured discussion. It aimed at identifying limitations, potentials, and general attitudes of the participants regarding the interaction design. To ease the discussion, it was structured into (i) collecting feedback on potentials and weaknesses associated with the prototype system, (ii) additional features participant missed and would desire in a future system, (iii) use cases of how such an ambient display could be used in other application domains than a meeting room, and (iv) an open discussion on anything that we may have missed asking for. The workshop took exactly one hour.

4.3 Workshop Results and Discussion

Overall, four participants stated that they would want to have such a system installed in their offices. Only one participant stated that they had already established a daily routine of opening the office windows, which they believed fulfilled their needs. Most of the participants comments addressed potentials of the system. The main weakness and limitation of the system was identified as usability issues, but only if one would use the ambient lights to display more and diverse information. For example displaying a history of air quality for a single room or use animation on LED stripes to indicate bad air quality in another room.

The main insights gained from the workshop are centered around using technology for ambient interaction to enabling a new form of “ambient materiality“, which grants users more agency associated with air quality in future urban spaces, which will be discussed hereafter.

Towards Interacting with “Ambient Materiality”. We summarize a set of comments made by the participants with the topic “Ambient Materiality”, because comments highlighted how the ambient and located light sources provided “material qualities” to air, which is usually invisible, non-graspable and nearly immaterial. Participants argued, for example, how by placing the LED stripes in different locations in the room (or an apartment) one might get a system being able to make recommendations on which windows to open and for how long to get fresh air to desired spaces (and rooms). Furthermore, participants highlighted that notifications on their mobiles would be very desirable and helpful in case one forgets to close a window or to receive reminders considering air

qualities in other/remote rooms. One participant provided as an example, how they don't open the windows in their bedroom at home because they usually forget to close the windows and end up freezing at nighttime. They argued that a reminder to close the windows once the room has established good air quality would be very much appreciated. However, participants also argued how there has to be balance between air quality, room temperature, humidity, and pollen. Consequently, there was a discussion about potential sensors also placed outside the windows to get recommendations on when to open windows to get good air quality without making it accidentally worse.

Digital Air Displays in Urban Spaces. A second topic that was discussed dealt with how digital air quality displays could transform various (urban) spaces. Workshop participants mentioned for example, how displays on public transportation could be useful in deciding which route or vehicle to choose for transportation. One participant argued that they didn't see a concrete benefit of having air quality displays in public spaces other than to sensitize fellow townsmen. One participant argued that there are already temperature displays which they appreciate without concrete benefit. Participants also discussed the potential of standardizing air quality sensors/displays on windows to monitor and regulate air quality in future digital cities.

In sum, the workshop has demonstrated that an ambient air quality display is very desirable and not only for meeting rooms but in general for urban regions. The participants' discussions have shed light on the diverse potential of ambient air quality displays and provided many ideas and inspirations for follow research and development work.

5 Conclusion

We presented the design and implementation of an ambient light notification system that unobtrusively informs people about the indoor air quality with the aim of improving it when necessary. Long-term measurements of our evaluation (without activated ambient lights) showed that the indoor air often became very poor without the users ventilating the room. Also, the ventilation phases often did not last long enough to improve the indoor air sufficiently. This part of the evaluation showed that an optimization of the ventilation behavior would be necessary or at least meaningful and effective. A workshop in the room (with activated ambient lights) showed that people understood the ambient notifications and would potentially lead to better indoor air quality. The workshop further revealed many more use cases such as including room temperature and humidity, weather, and outdoor air quality data into the semantics of ambient notifications. Even harmful gases outside the room, e.g., nitrogen dioxide or fine dust such as used in [17], may be considered as warnings in the ambient notifications in order to optimize ventilation phases. Additionally, our prototype could be researched in AAL settings by adding it to recommender systems like presented in [31].

Overall, we have shown that our system has the potential to improve indoor air quality and to create a healthier environment.

Acknowledgments. We want to thank Andrés Caro Quintal for his help with parts of the prototype.

References

1. <https://www.idevices.de/english/>
2. <https://www.sensirion.com/de/umweltsensoren/kohlendioxidsensoren-co2/>
3. https://www.bosch-sensortec.com/bst/products/all_products/bmp280
4. <https://www.tfa-dostmann.de/en/produkt/co2-monitor-airco2ntrol-mini/>
5. <https://ams.com/iam>
6. <https://www.eq-3.com/products/homematic.html>
7. <https://www.fhem.de/>
8. <https://www.drools.org/>
9. <https://www.letscontrolit.com/wiki/index.php?title=ESPEasy>
10. Gesundheitliche Bewertung von Kohlendioxid in der Innenraumluft. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz. **51**(11), 1358–1369 (2008). <https://link.springer.com/article/10.1007/s00103-008-0707-2>
11. Allen, J.G., et al.: Airplane pilot flight performance on 21 maneuvers in a flight simulator under varying carbon dioxide concentrations. *J. Expo. Sci. Environ. Epidemiol.* **29**(4), 457–468 (2018)
12. Allen, J.G., MacNaughton, P., Satish, U., Santanam, S., Vallarino, J., Spengler, J.D.: Associations of cognitive function scores with carbon dioxide, ventilation, and volatile organic compound exposures in office workers: a controlled exposure study of green and conventional office. *Environments* **124**(6), 805 (2016)
13. Azuma, K., Kagi, N., Yanagi, U., Osawa, H.: Effects of low-level inhalation exposure to carbon dioxide in indoor environments: a short review on human health and psychomotor performance. *Environ. Int.* **121**, 51–56 (2018)
14. Consolvo, S., Towle, J.: Evaluating an ambient display for the home. In: CHI 2005 Extended Abstracts, pp. 1304–1307. ACM (2005)
15. Dang, C.T., André, E.: A framework for the development of multi-display environment applications supporting interactive real-time portals. In: EICS 2014, pp. 45–54. ACM, New York (2014). <https://doi.org/10.1145/2607023.2607038>
16. Dang, C.T., Andre, E.: Acceptance of autonomy and cloud in the smart home and concerns. In: Dachsel, R., Weber, G. (eds.) *Mensch und Computer 2018 - Tagungsband*. Gesellschaft fuer Informatik e.V, Bonn (2018)
17. Dang, C.T., Seiderer, A., André, E.: Theodor: a step towards smart home applications with electronic noses. In: Proceedings of the 5th International Workshop on Sensor-based Activity Recognition and Interaction, iWOAR 2018, pp. 11:1–11:7. ACM, New York (2018)
18. De La Iglesia, D.H., De Paz, J.F., Villarrubia González, G., Barriuso, A.L., Bajo, J.: A context-aware indoor air quality system for sudden infant death syndrome prevention. *Sensors* **18**(3), 757 (2018)
19. Fortmann, J., Stratmann, T.C., Boll, S., Poppinga, B., Heuten, W.: Make me move at work! an ambient light display to increase physical activity. In: Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare, pp. 274–277 (2013)

20. Gunnarsen, L., Fanger, P.O.: Adaptation to indoor air pollution. *Environ. Int.* **18**(1), 43–54 (1992)
21. Mankoff, J., Dey, A.K., Hsieh, G., Kientz, J., Lederer, S., Ames, M.: Heuristic evaluation of ambient displays. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 169–176. ACM (2003)
22. Marques, G., Ferreira, C.R., Pitarma, R.: Indoor air quality assessment using a CO₂ monitoring system based on internet of things. *J. Med. Syst.* **43**(3), 67 (2019)
23. Mateevitsi, V., Reda, K., Leigh, J., Johnson, A.: The health bar: a persuasive ambient display to improve the office worker's well being. In: *Proceedings of the 5th Augmented Human International Conference*. ACM (2014)
24. Matviienko, A., et al.: Towards new ambient light systems: a close look at existing encodings of ambient light systems. *Interact. Des. Archit.* **2015**(26), 10–24 (2015)
25. Matviienko, A., et al.: Deriving design guidelines for ambient light systems. In: *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, pp. 267–277. ACM (2015)
26. Maula, H., Hongisto, V., Naatula, V., Haapakangas, A., Koskela, H.: The effect of low ventilation rate with elevated bioeffluent concentration on work performance, perceived indoor air quality, and health symptoms. *Indoor Air* **27**(6), 1141–1153 (2017)
27. Müller, H., Kazakova, A., Heuten, W., Boll, S.: Lighten up!—An ambient light progress bar using individually controllable LEDs. In: De Ruyter, B., Kameas, A., Chatzimisios, P., Mavrommati, I. (eds.) *AmI 2015*. LNCS, vol. 9425, pp. 109–124. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-26005-1_8
28. Müller, H., Kazakova, A., Pielot, M., Heuten, W., Boll, S.: Ambient Timer – unobtrusively reminding users of upcoming tasks with ambient light. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M. (eds.) *INTERACT 2013*. LNCS, vol. 8117, pp. 211–228. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40483-2_15
29. Occhialini, V., van Essen, H., Eggen, B.: Design and evaluation of an ambient display to support time management during meetings. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) *INTERACT 2011*. LNCS, vol. 6947, pp. 263–280. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23771-3_20
30. Ortiz Perez, A., Bierer, B., Scholz, L., Wöllenstein, J., Palzer, S.: A wireless gas sensor network to monitor indoor environmental quality in schools. *Sensors* **18**(12), 4345 (2018)
31. Rist, T., Seiderer, A., André, E.: Providing life-style-intervention to improve well-being of elderly people. In: Clua, E., Roque, L., Lugmayr, A., Tuomi, P. (eds.) *ICEC 2018*. LNCS, vol. 11112, pp. 362–367. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99426-0_45
32. Ritschel, H., Seiderer, A., Janowski, K., Aslan, I., André, E.: Drink-O-Mender: An adaptive robotic drink adviser. In: *Proceedings of the 3rd International Workshop on Multisensory Approaches to Human-Food Interaction, MHFI 2018*, pp. 3:1–3:8. ACM, New York (2018)
33. Satish, U., et al.: Is CO₂ an indoor pollutant? Direct effects of low-to-moderate CO₂ concentrations on human decision-making performance. *Environ. Health Perspect.* **120**(12), 1671 (2012)
34. Seiderer, A., Dang, C.T., André, E.: Exploring opportunistic ambient notifications in the smart home to enhance quality of live. In: Mokhtari, M., Abdulrazak, B., Aloulou, H. (eds.) *ICOST 2017*. LNCS, vol. 10461, pp. 151–160. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66188-9_13

35. Yu, B., Hu, J., Funk, M., Feijs, L.: Delight: biofeedback through ambient light for stress intervention and relaxation assistance. *Pers. Ubiquit. Comput.* **22**(4), 787–805 (2018)
36. Zhang, X., Wargocki, P., Lian, Z., Thyregod, C.: Effects of exposure to carbon dioxide and bioeffluents on perceived air quality, self-assessed acute health symptoms, and cognitive performance. *Indoor Air* **27**(1), 47–64 (2017)



ATHsENSE: An Experiment in Translating Urban Data to Multisensory Immersive Artistic Experiences in Public Space

Dimitris Charitos¹(✉), Iouliani Theona², Penny Papageorgopoulou¹,
Antonios Psaltis¹, Antonios Korosidis³, Dimitris Delinikolas¹,
Alexandros Drymonitis⁴, Natalia Arsenopoulou¹,
and Charalampos Rizopoulos¹

¹ National and Kapodistrian University of Athens, Athens, Greece
vedesign@otenet.gr

² National Technical University of Athens, Athens, Greece

³ University of Macedonia, Thessaloniki, Greece

⁴ Royal Birmingham Conservatoire, Birmingham City University,
Birmingham, UK

Abstract. This paper presents ATHsENSE, a multisensory interactive installation art project employing ubiquitous computing technologies, various sensors, virtual reality interfaces, multi-channel audio displays and interactive light structures. The project explores the concept of a smart city, through creatively translating the urban data produced by the environment of Athens and by its citizens. ATHsENSE is implemented through ubiquitous computing infrastructure, combining the use of sensors, location-based technology and mobile devices, in order to form a network of human and non-human sensors across the city. This structure enables the acquisition of rich urban data, which are eventually creatively translated in different artistic multisensory representations, displayed in physical and virtual space.

Keywords: Urban data · Environmental sensing · Data visualization · Virtual reality · Data sonification · Citizen sensing

1 Introduction

ATHsENSE is a multisensory digital media installation, comprising four interactive artworks, employing ubiquitous computing technologies, various sensors, virtual reality interfaces, multi-channel audio displays and interactive light structures, which aims to explore novel ways for (re)presenting and experiencing city related data. ATHsENSE was conceptualised by the Spatial Media Research Group¹, a multidisciplinary group of artists and scientists and its implementation was funded by the Municipality of Athens in the context of the “Interventions in the City” project, under the Polis2 pilot programme, which aimed to “...create and showcase the city’s

¹ <https://spatialmedia.ntlab.gr/>.

*identity*². ATHsENSE was submitted to the strand “Serafio – Public building and digital art”. Serafio is a sports, culture and innovation complex that is open all days of the week and accessible to all citizens. The call aimed for an artistic intervention that would employ digital technologies and exploit urban data in order to reinforce this particular character of the complex as an open creative hub, to demonstrate its role, position and relation with(in) the urban context, while at the same time, creating a unique experience for its visitors. A designated jury evaluated all of the proposals and selected ATHsENSE for development. The artwork will be installed in 2019 at Serafio and will remain exhibited there for a year. ATHsENSE will support interaction not just with the visitors of the Serafio complex but also with citizens that wish to experience aspects of it from afar.

This paper aims at presenting the ATHsENSE project and particularly focuses on the manner in which ubiquitous computing, virtual reality, physical computing, digital audio and sensing technologies are employed for the purposes of an artistic installation utilising urban data. Moreover, the deployment of the aforementioned technologies in complex systems and the process of exploring their expressive potential (via tweaking, tinkering, hacking) are also discussed.

2 State of the Art and Related Work

As the amount of generated data is exponentially increased in contemporary cities, emergent new urban conditions may transform the way in which citizens perceive these environments. Unstructured data eventually become structured, collected in real time through mobile devices, sensors, etc., processed and stored at low cost, in order to be used for different purposes pertaining to the technologically mediated urban infrastructures and sustainability. Different types of interfaces are employed to communicate the data and create new situated experiences, allowing for new artistic practices to emerge. Several related artworks which have already been exhibited, address the issues of data visibility, openness, usage and exploitation. Furthermore, a recurrent theme of artistic exploration has been the prospect of urban data fostering citizens’ participation (Verhoeff and Van Es 2018, p. 118). Next, we present some notable examples of such artworks.

Deep City (2015), a project by Ursula Feuersinger, created upon the façade of Ars Electronica Center (Linz, Austria) was produced with the purpose of exploring “the collective information that defines a city’s present and future” (Ars Electronica, 2015). Having collected user-generated data from four cities (Linz, Vienna, Berlin, New York), the artist investigated the occurring relationship strains between citizens, spaces and available resources, revealing at the same time publicly, certain invisible layers of data. The data were visualized with the use of artifacts of different light colour on the walls of the main building, transforming it into a visual timeline of urban data. The installation encouraged citizen participation through bodily interaction and accentuated the significance of the building at the urban space as a landmark.

² <http://www.polis2.thisisathens.org/en/>.

Similarly, *Lightswarm* (2017), created by Future Cities Lab highlighted the hybrid urban landscape as comprised by the physical space and its digital counterparts, and more specifically the sonic real time data of the surrounding space, captured by sound sensors. *Lightswarm* was implemented on the façade of Yerba Buena Center for the Arts (San Francisco, USA), where, “swarms” of light represented the noise levels of the area in real time, while vibration sensors turn the building into a “real-time urban sensor, an instrument to constantly sense the city” (Gattegno and Johnson 2017). Future Labs also created the on-site installation named *Murmur Wall* (2017) at the Yerba Buena Gardens (San Francisco, USA) using LED light and digital text displays, in order to visualize the online data shared by the users of the area while online. By monitoring data deriving from different search engines, the installation could be considered as a public sphere where the citizens could see in real time the keywords searched by their fellow citizens, enacting discussions at the physical space among the visitors. *Lightweave* (2017), developed by Future Cities Lab as well, located at an underpass railway (San Francisco, USA), acted as a mediator of urban sound and light; sound and vibration sensors captured the noise of the environment, which were then translated into dynamic light patterns, implemented with the use of woven lights across the railway.

This Is Where We Are (2016) by i-Dat, exhibited at the Tate Modern Digital (UK), explored a combination of data, including behavioural (using movement sensors on site), social media (using hashtags) and environmental (using sensors) data, translated into the movement, shape and color of algorithmic sculptures. Therefore, the audience members were invited to interact with the sculptures at the hybrid exhibition space, eventually also engaging in social interaction, since the sculptures responded to the collective movement of the audience.

The Nemesis Machine – From Metropolis to Megalopolis to Ecumenopolis (2019) by Stanza is a multilayered installation, which is not site-specific and has been exhibited to a number of countries. The artwork explores different themes pertaining to the ownership of the data and the urban space, as it responds to the real time data it receives from a wireless sensor network from the city and user-generated data from the web. *The Nemesis Machine* is a miniature mechanical metropolis where data are visualized with emphasis to their interactions and continuously evolves according to the received real time data stream. Moreover, the audience can interact with the artwork in real time.

Other notable installations include *Every Thing Every Time* (2018/9) by Naho Matsuda (Newcastle, UK/Austin, Texas, USA), which translates the daily stream of captured data (environmental, traffic, travel, etc.) from a network of sensors to a poem displayed at mechanical split-flap boards across the city and Particle Falls (2008–2018) by Andrea Polli, which translated the air pollution data to streams of coloured pixels, projected on the cities’ buildings.

3 Concept and Description of the Installation

The main objective of ATHsENSE is to creatively experiment and investigate various forms of innovative urban data representations, in order to create an ambience, an environmental composition, in the area surrounding the premises of Serafio, that translates, not easily discernible, environmental conditions into multisensory stimuli, affording an engaging multisensory experience for the visitors of the building.

Two methods are used primarily to gather the data:

1. Arduino based sensor kits collect data pertaining to different environmental parameters. In particular, they measure temperature, luminosity, humidity, as well as noise level and atmospheric pollution (Carbon Monoxide CO and Dioxide CO₂, Benzene and solid particles). One of the sensor kits is located at Serafio, whereas the others are dispersed throughout the city of Athens. The sensors measure the fluctuating values of the aforementioned set of parameters regularly during the day, seven days a week.
2. Citizens and visitors of Athens may also provide data regarding the way they feel about the urban environment they are located in, by responding to a simple, brief questionnaire available in the form of a web-based application, accessible either by a smartphone or a personal computer. The data gathered pertains to the visual, auditory and olfactory perception of environmental parameters from individuals on the move around the city of Athens, and in particular to their emotional response to them. Should the users of the application consent, their data input can be geolocated, otherwise user generated data is processed as a general indication of the state of citizens throughout the day.

ATHsENSE unfolds simultaneously in multiple spatio-technical levels: (a). the infrastructural level, (b). the physical level, (c). the virtual level.

The *infrastructural level* remains largely unseen to the visitor of the artwork. It essentially constitutes the system architecture performing in the background processes of data collection, processing and translation. The system architecture is further organized in two subsystems: (a) the data collection and processing subsystem dispersed throughout Athens and (b). the application subsystem, located mainly in the premises of the Serafio complex. This structure is presented in detail in the next section of the paper.

The *physical level* consists of the technical components of that manifest the output of the data translation in various forms of representation on location, in the surroundings of the Serafio complex. These components include displays, physical objects and custom-made structures communicating audio, visual and kinaesthetic stimuli. They are appropriately arranged and embedded into the spatial context of the building and the urban environment, bearing much of the aesthetic qualities of the artwork and affording a hybrid spatial experience.

Finally, the artwork constantly evolves at a *virtual level*. All types of captured data are translated, visualized and projected in real time on a wall adjacent to the main entrance of Serafio, in an artistic composition, which establishes inherent connections between the physical and virtual environment. This visual representation transforms



Fig. 1. View of the virtual environment visualization of all data sets

continuously, according to the measurements of the environmental conditions and the input received by the citizens.

The elements of physical and virtual level are part of the application subsystem. Their output is organised in the four site specific installations that together constitute the ATHsENSE artwork: (a) a large scale projection of the virtual environment, (b) a sound installation consisting of a 6 channel soundscape and an assemblage of kinetic objects making acoustic sounds, (c). a spatial composition of 8 LED matrix displays, (d). An interactive light structure. These installations afford varying degrees of interactivity for the purposes of providing to the visitors an engaging and meaningful experience:

- Citizens can actively contribute to the data collection by responding to the ATHsENSE web-based application. Subsequently, their feedback is integrated in the overall artistic composition as either a part of the visualisation (see above installation a), or the data sonification (installation b), In particular, the textual input is projected on the LED matrix displays (installation c). As these displays are placed in the main patio of the complex, an area that serves as a focal meeting point where groups of people are consistently gathered, the latter affordance is regarded as a playful element to encourage social interactions and spontaneous communication among members of the audience, mediated by the artwork.
- Visitors of the complex can also engage in embodied interactions with specific elements of the artwork. In what concerns the light structure (see above, installation d) the audience can manipulate the intensity of certain lights by the use of their voice (producing loud voices or soothing sounds) while the light intensity of the rest of the light composition relates with the environmental noise levels.

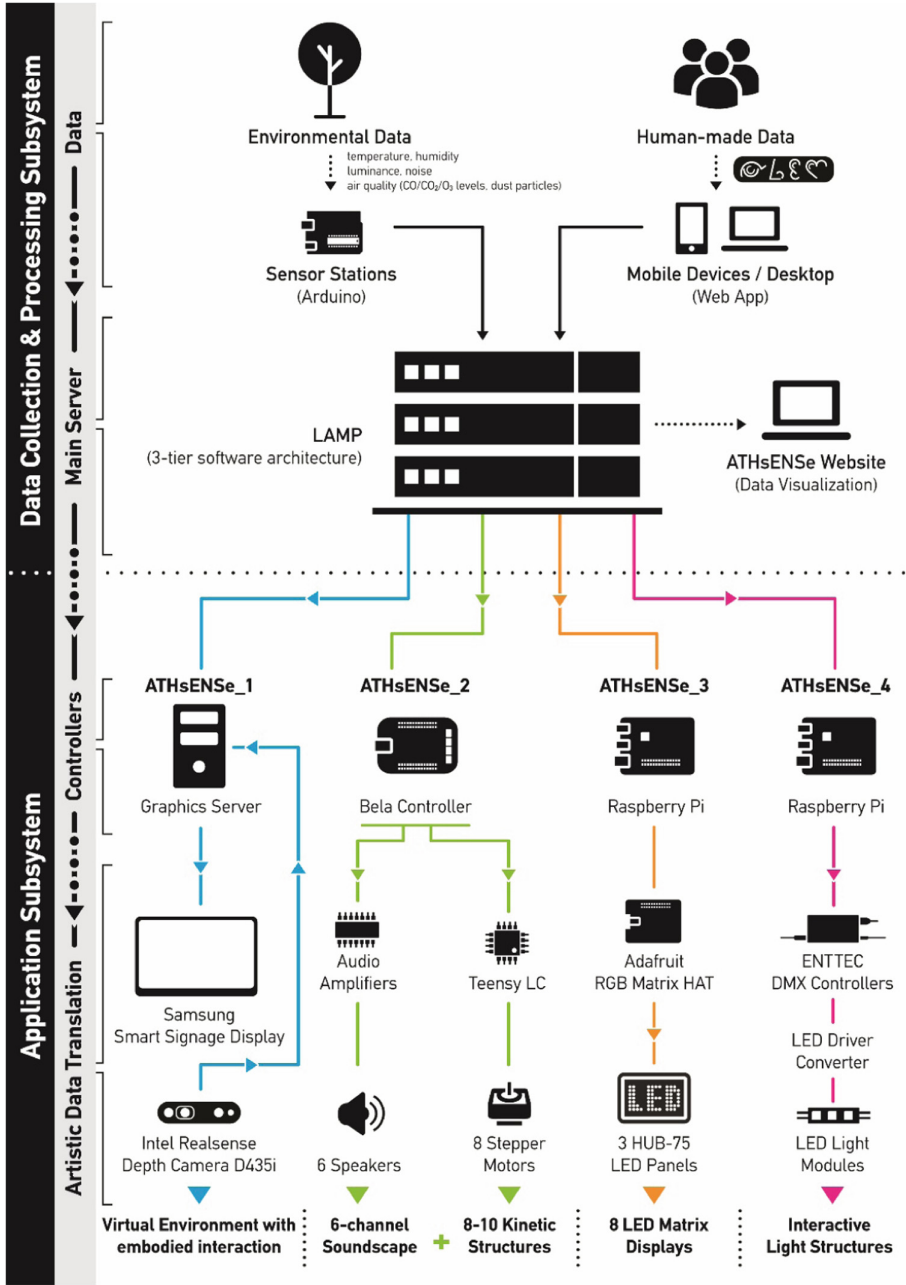


Fig. 2. Schematic of ATHsENSE system architecture

- Finally, the audience of the artwork can interact with certain components of the virtual environment, through bodily gestures and movement, resulting into a smooth, yet continuous transformation of the overall data visualisation.

In the next sections of the paper the technological infrastructure that enables the interactive artwork ATHsENSE to function is described in detail.

4 System Architecture

The ATHsENSE system architecture is shown in Fig. 2. Essentially the components that the ATHsENSE system consists of have been categorized in two main subsystems, namely (1) the *data collection and processing subsystem* and (2) the *application subsystem*. Low level processing of urban data takes place at the former level and higher level processing and translation of urban data into multisensory representations and interactive experiences are taking place at the latter level.

The data collection and processing subsystem consists of three components:

- the main server,
- the sensor stations and
- any remotely located mobile or desktop personal devices that may provide input to the system.

The application subsystem consists of:

- the graphics display server and its client used for the visualization of the data on the large visual display,
- eight LED matrix displays and their controllers,
- a set of lights and their controllers,
- a microphone, and a set of speakers
- a 6-channel audio display and
- an assemblage of kinetic objects consisting of step motors and their controllers.

These components are further described in the following section.

4.1 Data Collection and Processing Subsystem

Main Server. The main server of the system (web server) has been implemented in order to perform collection, processing and communication tasks between the web application producing user-generated and location data, automated environmental sensor data feeds and the application subsystem. The implemented three tier architecture makes this server the most vital component of the system, since it handles the communication between the front end (web application), middle tier and back end.

Furthermore, the server integrates the supervisory module (SM), which includes essential components for system initialization and configuration settings. An equally important part of the SM is also a high level data control and filtering algorithm for

user-generated textual data before they are forwarded for public display at the LED panels. The aforementioned module was developed in Python and implemented in Linux platform with GNU licensed software.

The server has been installed with Ubuntu Server 18.04.2 LTS, LAMP (Linux, Apache, MySQL and PHP) and OpenSSH. The collected data from the web application and the sensor stations are stored in databases at the back end with the use of MySQL database server. Each type of data is consequently exported in JSON format in order to be used by the application subsystem. Furthermore, Arduino IDE, installed at the server is used for the monitoring of all sensors, as well as the collection and scaling of the data captured from these devices.

Sensor Stations. The environmental data are collected by two sensor stations, consisting of two sets of sensors integrated in two Arduino Mega boards respectively, placed outdoors, at two different regions of the centre of Athens. The sensors collect data pertaining to temperature, humidity, luminance, CO/CO₂ levels, noise, dust particles, and benzene. The stations use autonomous power supply and transmit the captured data to the main server through an ethernet connection. More specifically, the stations consist of the following hardware:

- Arduino Mega 2560 Board (with ethernet shield)
- Waveshare MQ-7 Gas Sensor
- Waveshare DHT11 Temperature-Humidity Sensor
- Waveshare Light Sensor
- Waveshare Sound Sensor
- Waveshare Dust Sensor
- AMS CCS811 Gas Sensor
- Waveshare MQ-135 Gas Sensor

Desktop or Mobile Personal Devices. The user-generated data (ordinal and textual) are generated through the use of a web application in the form of a questionnaire, which citizens may access through a standard web browser with their desktop or mobile devices. Furthermore, the application is location aware and allows for the consensual provision of location data by citizens. The software architecture of the web application follows a three tier paradigm, hence the front end is implemented with the use of standard web technologies, including HTML5, CSS3, Javascript, Angular and Bootstrap Framework.

4.2 Application Subsystem

Graphics Display Server-and Visual Display. The graphics display server is a core component of the application subsystem since it handles the visualization of all types of data in a composite virtual world depicting all the streams of the collected data in real time, which eventually are presented at a large scale visual display, located at the entrance of the building. The visualization is created with the use of Unity software and

processed with an Nvidia GeForce RTX 2080 8 GB graphics card. As depicted in Fig. 1, the elements of the visualization correspond to different environmental parameters, as well as user-generated data. Embodied interaction with visitors on site is supported by the use of an Intel RealSense i435 depth camera. This input device captures the movement of visitors' bodies located in front of the visual display and triggers the transformation of form, position and orientation of elements of the visualization, projected onto the display.

LED Panels. Citizens may provide text input, up to 180 characters long, through their mobile or desktop devices, thus communicating their emotional state about the environment they experience, there and then. Data may also include information about the location of the sender if they give their consent. These text messages are fed through to the web-server for screening and conformance testing. Messages are then fed from the web-server via Wi-Fi through to each one of 8 LED matrix displays. Display priority is optimised so that preference is given to people located on site so that they may observe their messages displayed on the LED panels, a few seconds after they input them. The LED panel sub-system including the hardware, enclosures, controllers (Raspberry Pi 3 and Adafruit RGB Matrix HAT + RTC) and software development for communication and functionality has been implemented in Python IDE under Linux platform.

5 The ATHsENSE Content and Experience as a Multisensory Translation of Urban Data

5.1 Data Visualization

In recent years, data visualization has emerged as a novel form of artistic practice, crossing the strict boundaries of scientific representation of data. The rapid evolution of computer graphics hardware and software, as well as the constant increase of open data sources has allowed for the advancement of data visualization with creative elements that allow for immediate artistic and social impact. According to Viégas and Wattenberg (2007, p. 184), what shifts data visualization to artistic visualization is the intention of the creator to make art instead of serving mere scientific purposes; an artistic visualization comprises of two elements, namely the use of actual data in order to create an artwork and the subjective character of an aesthetically pleasing visualization towards an artistic result. Art can be ugly, odd or unpleasant, yet it may convey an important message leading to different amounts of aesthetic appreciation, interpreting audience's aesthetic pleasure or detachment (Sheppard 1987, p. 65).

A central element of ATHsENSE is the composite virtual world depicting all the streams of collected data in real time, visualised in an artistic interpretation, which aims to make the audience perceive the Athenian environment in different ways that expand beyond daily sensory experience, as well as the citizens' emotional response to it. By creating this metaphorical representation of data, the Spatial Media Research Group seek to initiate a dialogue between the audience members in the public space of Serafio, as well as between the digital visitors of the artwork, in the virtual public space.

The artists' group have also intended to keep a balance between a multisensory symbolic representation of the data, aiming at communicating their characteristics and meaning at a denotative level and a more expressive, connotative and possibly deconstructive interpretation of these data, which is more in accordance with a visualization as a work of art and not a descriptive presentation of the data sets. As depicted in Fig. 1, the elements of the visualization, corresponding to different environmental parameters, give a visual form to commonly imperceptible environmental hazards, which tend to be ignored by citizens in their everyday lives. The composite, a visual narrative of the Athenian environmental conditions is designed to raise awareness towards city's pollution and provide a ground for a public discussion concerning the issue.

The emotional response of each citizen using the ATHsENSE web application is represented as a floating polygonal shape,

- the colour of this shape signifies the citizen's affective state with regards to the visual stimuli they experience where they are
- the movement/animation of this shape signifies the citizen's affective state with regards to the auditory stimuli they experience where they are
- the level of opaqueness/transparency of this polygon's form signifies the citizen's affective state with regards to the olfactory stimuli they experience where they are
- ATHsENSE makes use of the above user-generated data in relation to the geographical location of the citizens that provide it if they consent to this location being captured. This location is consequently mapped to the bottom surface of the visualization and thus determines the x and z coordinates of the position of the polygon in the representational context.

5.2 Data Translation to Sound and Kinetic Artwork

The audio installation ATHsENSE_2 comprises of a micro-computer doing HTTP requests to the server of the project, in order to receive data from the sensors attached to the server, and map the results of the request to various auditory processes. Along with the electronic sounds created by the computer, there is an accompanying section comprising assemblages of found objects being excited by step motors thus becoming kinetic objects which dynamically generate acoustic sounds. The cross-fade between the electronic and acoustic sounds depends on the sensor data and thresholds set on these values. The micro-computer determines when these thresholds have been crossed either direction and triggers the electronic or acoustic processes respectively.

The electronic sounds used for the installation are not a precise sonification of the sensor data, but an artistic interpretation of them. Depending on the type of data, layers of different procedural sounds fade in and out. The data used are based on the values of CO, CO₂, dust particles, humidity, light, temperature, and other similar types of data. The light data determines the overall sound volume of the installation, and it is the main value that sets whether the electronic or the acoustic processes are triggered. Light amount is also translated to audible noise, with a white noise source being filtered in various ways, where, the more light the more frequencies pass through the filters, resulting in a more noisy and loud environment, and the less light, the less frequencies

pass, resulting in a more musical and melodic soundscape. Other data are mapped to different soundscapes depending on their amount and impact on the environment. The resulting sound environment is presented through a 6-channel audio display (3 amplifiers and 3 pairs of speakers).

The programming environment Pure Data was used for the realisation of this installation, which runs on a Bela micro-computer. The HTTP requests are being done from within the Pure Data patch, via the rest object from the PuRestJson external library, compiled for the architecture of the Bela.

5.3 LED MATRIX Displays

This part of the ATHsENSE project consists of a number of LED matrix displays which are placed on the central wall located on the main patio of Serafio. The displays present textual content from messages posted by citizens that is drawn from the project's central server.

In the urban environment, LED matrix displays are widely used in a variety of applications, from signage systems and transportation timetables to advertising or stock exchange tables. The use of LED matrix displays on artistic installations for the projection of textual messages is not something new or radical. Artists have been experimenting with the specific medium since the 1980's. Jenny Holzer, one of the pioneers of led art, commenting on her artworks on an interview with Kiki Smith stated that she 'uses language because she wanted to offer content that people - not necessarily art people - could understand' (Holzer 2012). The textual content of the displays in ATHsENSE_3 however, is created by and for the people, in an interactive two-way communication manner.

Through the previously described in [3.1] ATHsENSE web application, citizens are invited to comment on their overall urban experience, their thoughts and reflections on the city. Thus, the artwork serves as a communication channel between the people and the city, a medium for social interaction, environmental awareness and an incentive for action-taking. The wall of Serafio building is turned into an interface, an augmented space (Manovich 2006, p. 223) where information drawn from the sensors and messages delivered from the users coexist. The simultaneous projection of objective information and textual messages opens a dialogue between the city and the people, creating a kind of non-linear urban narrative.

5.4 Data Translation to Sound and Light Artwork

“... we claim that what a pixel-less display might lose in resolution and information bandwidth, it could make up in a richer, more intriguing and memorable experience that nonetheless communicates information and insight (Moere 2008).

In the southern edge of Serafio, across the parking area there is a pavilion, which sits quite remotely as it is not immediately connected to the sport activities of the main buildings (Fig. 3). It presents itself as a potential space for contemplation and introspection. In this area the artists proposed the creation of ATHsENSE_4, a minimal interactive artwork consisting of light and sound.



Fig. 3. View of the light and sound installation at the Serafio pavillion

The urban data that drives this light installation is noise pollution. It is often the case that strong sounds are followed by strong light. This stands true for many natural phenomena such as thunder and lightning, or explosions, but also for man-made sounds like sirens, or live shows. Both strong sound and light attract our attention and make a strong physical impact. However in the case of this installation the opposite connection is made. The louder the sound is the lower the light becomes. As a visual metaphor of noise pollution oppressing and strangling us in cities, the light appears to be fading depending on the strength of noise that it is connected to.

A central light structure, comprises 6 long metal boxes within which LED lights are positioned. The intensity of light projected by this light structure corresponds to the noise of the city. Additionally, a speaker lets the bystanders hear the noise that is recorded so that they can understand where the fluctuations of light originate from. Alongside the main light that corresponds to the city's noise pollution there is also one longer light, mounted above a centrally placed bench at the periphery of the pavillion. This light responds to live sound input provided by the visitors of the installation and captured by a microphone, positioned on top of the bench. When people sitting at each bench are producing louder noises the corresponding light fades in the same way that the main light fades from the city's noise.

The light installation ATHsENSE_4 runs on a Raspberry Pi micro-computer, which runs a Pure Data patch. The patch makes HTTP requests via the rest object from the PuRestJson library, and collects information about the noise levels. These values are then transferred to an EntTec DMX controller via the com port external Pd object, which in turn controls the lights.

6 Epilogue and Future Work

ATHsENSE is a multisensory interactive media installation, employing ubiquitous computing technologies, various sensors, virtual reality interfaces, multi-channel audio displays and interactive light structures, ultimately aiming at investigating urban data as material for expressive purposes. It has not aimed to extend the current state of the art at a purely technological level. The innovative aspects of this artwork lie mostly in that:

- the above mentioned technologies are employed for creating a complex and engaging multisensory artistic experience
- the various elements of the artwork (physical objects, hardware, interfaces, visualization and other multisensory forms of representation) have been designed in terms of form and content so as to present an aesthetically coherent and meaningful overall composition.


ATHsENSE will be installed at the Serafio complex in Athens in October 2019. Following its installation, ATHsENSE will remain there for at least one year. The Spatial Media Research Group have planned a series of qualitative and quantitative studies in order to evaluate the experience afforded by the artwork and its impact at a social level.

References

- Verhoeff, N., van Es, K.F.: Situated Installations for Urban Data Visualization: Interfacing the Archive-City, pp. 117–136 (2018)
- Ars Electronica: Connecting Cities: Deep City. [online] Post City (2015). <https://ars.electronica.art/postcity/en/deep-city/>. Accessed 15 Sept 2019
- Gattegno, N., Johnson, J.K.: Computational Doppelgängers. *Technol.| Archit. + Des.* **1**(2), 146–149 (2017)
- Holzer, J.: Interview by Kiki Smith (2012). <https://www.interviewmagazine.com/art/jenny-holzer>. Accessed 27 Apr 2019
- Manovich, L.: The poetics of augmented space. *Vis. Commun.* **5**(2), 219–240 (2006). <https://doi.org/10.1177/1470357206065527>
- Moere, A.V.: Beyond the tyranny of the pixel: exploring the physicality of information visualization. In: 2008 12th International Conference Information Visualisation, pp. 469–474. IEEE (2008)
- Sheppard, A.D.: *Aesthetics: An Introduction to the Philosophy of Art*. Oxford University Press on Demand (1987)
- Viégas, F.B., Wattenberg, M.: Artistic data visualization: beyond visual analytics. In: Schuler, D. (ed.) OCSC 2007. LNCS, vol. 4564, pp. 182–191. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-73257-0_21



Characterization of Individual Mobility for Non-routine Scenarios from Crowd Sensing and Clustered Data

Inês Cunha¹, João Simões², Ana Alves^{2,3} , Rui Gomes³,
and Anabela Ribeiro¹

¹ Centre for Territory, Transports and Environment, University of Coimbra,
Polo II, 3030-788 Coimbra, Portugal

inesfariadacunha@gmail.com, anabela@dec.uc.pt

² Coimbra Institute of Engineering, Polytechnic Institute of Coimbra,
3030-199 Coimbra, Portugal
a21220126@alunos.isec.pt

³ Centre for Informatics and Systems, University of Coimbra, Polo II,
3030-290 Coimbra, Portugal
{ana, ruig}@dei.uc.pt

Abstract. Demand for leisure activities has increased due to some reasons such as increasing wealth, ageing populations and changing lifestyles, however, the efficiency of public transport system relies on solid demand levels and well-established mobility patterns and, so, providing quality public transportation is extremely expensive in low, variable and unpredictable demand scenarios, as it is the case of non-routine trips. Better prediction estimations about the trip purpose helps to anticipate the transport demand and consequently improve its planning. This paper addresses the contribution in comparing the traditional approach of considering municipality division to study such trips against a proposed approach based on clustering of dense concentration of services in the urban space. In our case, POIs (Points of Interest) collected from social networks (e.g. Foursquare) represent these services. These trips were associated with the territory using two different approaches: ‘municipalities’ and ‘clusters’ and then related with the likelihood of choosing a POI category (Points-of-Interest). The results obtained for both geographical approaches are then compared considering a multinomial model to check for differences in destination choice. The variables of distance travelled, travel time and whether the trip was made on a weekday or a weekend had a significant contribution in the choice of destination using municipalities approach. Using clusters approach, the results are similar but the accuracy is improved and due to more significant results to more categories of destinations, more conclusions can be drawn. These results lead us to believe that a cluster-based analysis using georeferenced data from social media can contribute significantly better than a territorial-based analysis to the study of non-routine mobility. We also contribute to the knowledge of patterns of this type of travel, a type of trips that is still poorly valued and difficult to study. Nevertheless, it would be worth a more extensive analysis, such as analysing more variables or even during a larger period.

Keywords: Urban mobility · Destination choice modelling · Clustering analysis

1 Introduction

Traditional transport supply design criteria require satisfying the highest demand hours, typically occurring during the work commute. However, in recent decades, the demand for leisure activities has increased due to some reasons such as increasing wealth, ageing populations and changing lifestyles, yet non-routine trips have only recently received more attention. Understanding personal travel patterns and modelling travel demand is essential to plan sustainable urban transportation systems to fulfil citizens' mobility needs. To do this effectively and timely, urban and transportation planners need a dynamic way to profile the movement of people and vehicles. Traditional survey methods are expensive, time consuming and give planners only a picture of what has happened.

In contrast, an emerging field of research uses the wide deployment of pervasive computing devices (cell phone, smart card, GPS devices and digital cameras) and transport system records (e.g. ticket validation counts; traffic counts) providing unprecedented digital footprints, telling where and when people are, for “urban sensing”. Moreover, the composition of social networks and human interactions is crucial not only for understanding social activities, but also for travel patterns. With millions of users around the world, transportation researchers have also realized the potential of Social Network Analysis (SNA) for travel demand modelling and analysis.

To date there is little published research, however, on how to realize this opportunity for the sector by capturing the views, needs and experiences of the travelling public in a timely and direct fashion through social media. All these mobility data, together with modern techniques for geo-processing, SNA, and data fusion, offer new possibilities for deriving activity destinations and allowing us to link cyber and physical activities through user interactions, in ways that can be usefully incorporated into models of land use and transportation interactions.

With data registered in mobile phones by users, using an App designated by SenseMyFEUP¹ [1], it was possible to characterize some aspects of the mobility of these users, as the destination selected for non-routine trips. The selection of this type of trips is related with the importance of recording some patterns that are not captured by the most usual mobility studies and that are important to characterize, in order to identify the gaps and to adapt the public transport offer, trying to decrease the use of the car, specially during the night-time. This type of information has been cross-referenced with social network data that records entries and preferences in places with night activities to characterize these patterns as accurately as possible.

It was possible to see that most of the trips that were made had origin and destination in the same municipality, with a greater tendency on weekdays. During weekends, the tendency to move to other municipalities increases, pointing to the availability

¹ <https://sensemycity.up.pt/project/sensemyfeup/> (September 2019).

to travel more distances being higher on weekends compared to weekdays. In the Municipality of Porto, Portugal, there was a greater number of movements in parishes with a higher offer of events/POIs, showing the attractiveness created by them. Geographic proximity also showed an influence, demonstrated both by the low travel times and by the proportion of trips to neighbouring municipalities concerning remote municipalities.

Regarding the modes of transport used during non-routine hours, the car was the most used, confirming the great tendency to use this mode already presented in other previous studies based on regular trips in the Porto Metropolitan Area, followed by foot. It was also identified the importance that certain variables had in choosing the destination, using a multinomial logit regression. The variables of distance travelled the number of check-ins, travel mode and whether the trip was made on a weekday or a weekend had a significant contribution.

The collected dataset of trips and the results will be discussed further, and although allowing very interesting results, and in order to obtain more reliable results, we have tried a new approach in this work using clustering analysis. With this, we identified agglomeration of POIs, i.e., the services that are offered in a concentrated way in the urban space. And, on the other hand, we also took into account the concentration of events that usually happen in certain areas of the city, i.e., looking for social events that are not related specifically to the importance of their venues but which are important enough to attract people. With this in mind, clusters of important areas were computed to represent the origin or the destination of a given trip, rather than the traditional municipality used in most of studies. The results obtained for both geographical approaches are then compared, considering the multinomial model studied to check which approach gives the best results for non-routine travel analysis.

In order to promote a broader discussion about the influence of geographical units in the analysis of mobility patterns and spatial data clustering, we present a review of the literature in Sect. 2. In Sect. 3, we present the proposed approaches of the comparison study (territorial units versus cluster techniques). In Sect. 4, we present the results and some discussion about them. Finally, on Sect. 5, conclusions are drawn and next steps are proposed pointing the main inputs for future research.

2 Literature Review

2.1 Studies on Non-routine Mobility Patterns

In recent decades, the demand for leisure activities has increased due to some reasons such as increasing wealth, ageing populations and changing lifestyles. This increase in demand has led to higher levels of emissions and congestion and, therefore, to greater concern in research [2].

Daily travel has already been comprehensively studied. However, non-routine trips have only recently received more attention, despite their environmental impact. Unlike fixed trips (i.e., trips to work where the arrival time and the trip destination are pre-defined), leisure activities offer more optimization flexibility since the activity

destination and the arrival times of individuals can vary [3]. The study of these trips is the focus of this paper.

An important issue to realize in this type of travel is that the option of starting this type of activity is confined to certain times of the day due to restrictions imposed by other activities [4]. Regarding the frequency of leisure trips per week, Tarigan and Kitamura [5] realized that the number of cars at home has a positive impact on the number of activities per week related to shopping and recreational activities, but not so much in social contact activities. On the other hand, the more bicycles there are at home, the more trips are per week related to sports and nature activities. Individuals living in the suburbs tended less to participate in nature-related activities than those living in the city. They also found that factors such as gender, age, life-cycle stage, vehicle ownership, and residence location have a significant influence on leisure travel patterns.

Another study focusing on leisure trips characteristics was that carried out by Sánchez et al. [6]. They found that men perform more leisure trips regardless of age, women jaunt average distances, and also that leisure trips are made closer to home regardless of the size and type of town. A more specific study was carried out by Sener et al. [7] where they analysed physical leisure activities. They concluded that socio-economic, demographic, domestic, employment-related and environmental variables affect the choice of physical recreation activity, relative to location, time of day, the day of the week and social context. Moreover, a study was conducted to understand whether there were leisure travel patterns differences between people living in an urban district and people living in a small city in a suburb. Große et al. [8] found out that the urban structure of a residential place influences the constitution of daily mobility styles and that there is a greater tendency to perform weekend trips and holidays by people living in an urban district.

2.2 Crowd Sensing for Mobility Characterization

In the field of data collection for transport studies, several studies on mobility patterns have already been done using data from mobile phones and social networks, confirming the various advantages as data providers. Some of these advantages were presented by Calabrese et al. [9], such as lower cost, larger sample size, higher update frequency, more spatial and temporal coverage and, also, provide unprecedented “digital footprints”. However, they have also identified some gaps, which could be filled by traditional methods. Examples of such gaps are the impossibility to access socio-economic and demographic characteristics due to privacy issues; the high probability that mobile phone users do not represent a random sample of the population tending to be biased, and the difficulty of using the database because of its format is not ready for modelling.

There are also some challenges in comparing different datasets, even if they are related. Anda et al. [10] pointed out that the main ones are the different harvest periods and different spatial units. However, they also have the advantage of using different human mobility sensors and complementary datasets, such as travel diary data, to complement the importance of the data fusion approach.

In the field of transport planning, one of the advantages of these digital footprints is helping to reduce underreporting of trips, which is a common phenomenon in travel surveys. Thomas et al. [11] showed this conclusion by carrying out a mobile application study in the Netherlands for one month, registering departure and arrival times, origins and destinations, and travel reasons for a group of 615 volunteers. To compare automatically detected and reported trips, volunteers also had to participate in a requested web-based recall survey and answered additional questions. Most trips were identified without any apparent defects in the length or duration of the trip, and modes of transport were classified correctly in more than 80% of these trips.

Digital footprints come also from the Location-Based Social Networking (LBSN) such as “Foursquare”, “Gowalla”, “Google Latitude” and “Facebook Places”. These social networks are an indispensable source of voluntary geographic information [12]. While providing data from different age groups and site categories, geo-tagged social network data can complement regular household survey data to validate findings of mobility and urban structure and to reveal new insights about nature and human behaviour [13].

2.3 Clustering Techniques Applied to Urban Space

In [14], the authors identify seven groups of data analysis techniques usually used to understand urban mobility: visualization, statistics, logical heuristics, graph theory, optimization, clustering and classification.

Regarding clustering, a study implemented in the PublicSense mobile crowd sensing framework [15] in the city of Xi’an, in Japan, proposes the use of complaints platform data to understand the current problems of the city (mobility, constructions, among others), and thus supporting the decision making with respect of management of city resources to solve problems. This study applies clustering, namely K-means, to identify those areas that contained more complaints, most related to mobility problems, and therefore require more attention. This same algorithm is used by Quadri et al. [16] to identify three (K) classes of places visited by each individual person from Call Detail Record data to identify non-routine trips: Mostly Visited Places, Occasionally Visited Places and Exceptionally Visited Places.

The NSE (National Science Experiment) project [14] aims to analyse and understand the mobility of young students, namely the determination of patterns in the means of transport used, through the exclusive collection of access points extracted from the Wi-Fi network. The identification of POIs and trips is performed through the application of the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm. In another study [17] this same algorithm was used to group visited GPS locations visited by elderly people according to different profiles of trips.

Gan et al. [18] propose a framework of identifying urban mobility patterns and urban dynamics from a spatiotemporal perspective and pointing out the linkages between mobility and land cover/land use (LCLU). For this, they used data from the AFC (Automatic Fare Collection) system of Nanjing metro and applied the K-means clustering algorithm to isolate seven types of stations (depending of they are used to return home, go to work, etc.). Next, a multinomial logit regression was estimated to explore the relationship between each distinctive cluster and local land use characteristics.

2.4 Comparative Analysis

After the collection and analysis of the mentioned studies, it was verified that some have already been done considering non-routine trips (leisure, social, tourism, etc.). However, as this type of studies has only recently been given more relevance, they still have some gaps, such as: dispersion in calculation methods, data sources, time intervals and scopes, the accuracy and completeness of the measurement are compromised by survey designs, socio-psychological variables are often not controlled, and more [19].

Complementary, the use of digital footprints presents many advantages for the characterization of the mobility compared to traditional methods. However, several studies have found that for mobility research, the best way to obtain a database with quality is to combine information from different available data sets, from traditional sources like mobility surveys, to more recent ones, like mobile phones data [10, 11, 13].

Such digital footprints represent large volume of geospatial and crowd sensed data that brings a new perspective of more up-to-date mobility studies. Unsupervised techniques, as clustering, can deal with this type and massive quantity of data, which are not labelled and are produced in real-time fashion. To the best of our knowledge, none of the scientific works found in the literature proposed a comparative study to the characterization of non-routine trips using territorial divisions against density-based clusters. Our study proposes to develop a methodology that contributes to the improvement of this knowledge. In the methodology proposed, we also exploit the use of silhouette score as an internal and external measure to identify the best configuration values for each density-based clustering analysed.

3 Proposed Approach – Methodology

As already mentioned, the area defined for the study corresponds to the city of Porto and its surroundings, more specifically the Porto Metropolitan Area. What led us to this choice was the fact that there is large-scale data on social networks, both at points of interest and events. To detect and collect movements of people, the SenseMyFEUP [1] app was used to collect trip data from volunteers in April 2016, targeting students and personnel of the Faculty of Engineering of the University of Porto. Participants were asked to install the corresponding application on their smartphones that automatically collected mobility data during trips and launched questionnaires enquiring participants on the used transportation mode.

Some characteristics of the resulting dataset, such as errors in raw data and user engagement analysis, were published in [20, 21]. This dataset was made available by their authors in a given Data Access Agreement, and after trip segmentation and anonymization. It comprises data from 239 unique participants from the Faculty of Engineering of Porto. In total, 23600 trips were gathered over 6300 h of commuting time. Since this is a study focused on non-routine movements, we studied every day of that month only from 7 p.m. to 7 a.m., except for the weekends for which the full day was analysed. After this filtering, we reached a total of 2246 non-routine trips. Since the study hours depend on the day in study, the sample was divided into two distinct groups: weekdays and weekend days.

Foursquare² is a social network with the main objective to register the users' visits to a Point of Interest (check-ins). The user can classify the place and add pictures or comments. This data source has many POIs and provides an Application Programming Interface (API) to collect data in an automated way. These POIs are classified in categories that forms a taxonomy with ten top-level categories: *Arts & Entertainment*, *College & University*, *Event*, *Food*, *Nightlife Spot*, *Outdoors & Recreation*, *Professional & Other Places*, *Residence*, *Shop & Service* and *Travel & Transport*.

Infoporto³ is a website that gathers Porto city events, encompassing a portal to disseminate information about the Porto Region. It allows the consultation and research of events to take place in that region. It has a strong connection with Facebook Events in the way that every event advertised has its link to this social network. This service is available free of charge to the public. We developed a script in python to collect events dated April 2016.

3.1 Global Approach

The method developed comprehends four phases:

1. Data collected from SenseMyFEUP dataset and from social networks (Foursquare and InfoPorto);
2. Clustering all events and POIs collected in two different layers considering their density.
3. Data processing, making two similar samples: (a) one based on territorial division (2934 trips); (b) one based on clusters techniques (2091 trips, after excluding those that are not having a cluster of POIs as destination), both for the month of April 2016 and considering the trips regarding leisure purpose (inferred by the closest popular POI top-level category) from the following Foursquare categories: *Arts & Entertainment*; *Event*; *Food*; *Nightlife Spot*, and *Outdoors & Recreation*;
4. Analysis and discussion of the obtained results by using multinomial logistic regression for the destination choice model estimation and comparing the data obtained with the two samples;

The modelling approach is aimed at identifying the most determining factors in the choice of a given location. To support this analysis, it was used the Statistical Package for the Social Sciences (SPSS) software⁴.

3.2 Territorial Division

OD matrices were obtained by crossing information from the trips and the events and POIs, the latter only considering those with more than 500 check-ins (the most popular ones). At this stage, the trips were treated at the scale of the segments. With an area of influence defined as a circle of radius equal to 500 m and centre in the places of the

² <https://developer.foursquare.com/> (July 2019).

³ <https://www.infoporto.pt/en> (July 2019).

⁴ <https://www.ibm.com/analytics/spss-statistics-software> (July 2019).

events/POIs, the determination of the destinations of the trips was made. Thus, in the case of these matrices, the origins correspond to the beginning of the trip and the destination to the municipality where the event or the POI took place, or vice versa.

3.3 Stages of Clustering

The methodology applied aimed to identify the main areas, which are most sought after in terms of non-routine destinations in weekdays and weekend days, and an attempt to infer the possible purposes of these trips, which led SenseMyFeup users to travel to a particular destination. It was intended that the implemented methodology could answer questions such as which areas of Porto are most sought after and at what hours, which correspond to the period outside the daily life, i.e. during the week after 18h30 to 07h00 of the next day and the weekend. Also, the clustering process could help understanding which areas of the metropolitan area offer more services and which can therefore be considered as regions of high attractiveness.

The first step was to select points of interest in the city of Porto, which could be related to the trips made by users. In addition, events in the city were also selected that may be related to possible trips as well.

In order to identify the regions with the highest concentrations of points of interest and events, we chose to use the clustering algorithms HDBSCAN (Hierarchical DBSCAN), DBSCAN and OPTICS (Ordering Points to Identify the Clustering Structure). One of the main difficulties in using these three algorithms is the correct definition of the initial configuration, i.e. the minimum number of points and epsilon most indicated for the case study. To do this, we chose to use validation to determine the best configuration. A common approximation is based on the use of validation indexes [22].

The clustering assessment can be divided into three main categories: internal, external and relative. The external approach is based on prior knowledge of the data. In turn, internal indexes are used to measure cluster quality without external information. Relative approximations are used to compare different groupings or clusters.

Considering these approximations, and since the manual visualization of each possible configuration would be very painful and we do not have any prior knowledge of the data, it was decided to follow the use of an index that works for the both measures that can help us to identify the best configurations for each algorithm (HDBSCAN*, DBSCAN and OPTICS): relative and internal measures. These algorithms were selected considering that we need to group spatial data but there was no clue about the number of clusters (a pre requisite to other clustering algorithms, such as K-means).

In order to assess the clustering phase, from the algorithms above, we used the silhouette score index [22]. This index is based on the principle of maximum internal cohesion and maximum separation between clusters, that is, it measures how close the objects are to each other, and belonging to the same cluster and how far apart they are from objects of other clusters. The silhouette score values range from -1 to 1 . Values less than 0.25 mean that no substantial structure has been found, values ranging from 0.26 to 0.50 means that the structure found is weak and may be artificial, values ranging from 0.51 to 0.70 mean that a reasonable structure was found. Finally, values above 0.7 mean that a strong structure was found.

In order to achieve a larger volume of samples, we chose to select the clustering configurations that maximize the assessment index, the silhouette score, but which may also have resulted in more than 30 clusters for statistical significance. For each layer (POIs and events), the OPTICS and DBSCAN algorithms were run with epsilon varying between 0.0005 and 0.0024, in intervals of 0.0001. For each epsilon value, the minimum point value varied from 4 to 170.

In the case of HDSBCAN*, since it does not use epsilon, but just the minimum of points, the algorithm was run for each layer with this configuration parameter from 4 to 170. Table 1 presents the results of the assessment for the event layer, as an example of evaluation performed to select the best algorithm and configuration, respectively.

Table 1. Configuration obtained in the event layer, in the weekend days.

Algorithm	DBSCAN	OPTICS	HDBSCAN
Minimum points	4	4	4
Epsilon	0.001	0.0012	N/A
Resulting number of clusters	20	16	16
Silhouette score	0.85	0.83	0.74

3.4 Model Generation - Destination Choice Models

These models tried to answer the following question: The user's role, the mode of transport used, weekday vs weekend day, travel time, distance travelled and the number of check-ins in the event/POI may influence the destination choice? And, how can the geographic approach influence these relations?

Two models were tested: one model considered trips starting and ending in a parish/municipality of Porto and the other considered a cluster as destination.

The destination choice models considered several types of variables, including both trip-related characteristics, event/POI attributes, and individual socio-demographics characteristics.

Trip-related characteristics explored in our specifications included nominal variables such as the transport mode used for the trip; and, whether the trip was on a weekday or on a weekend day; and scale variables: the trip travel time; and, the distance travelled.

Event/POI attributes included two variables: number of check-ins in the event/POI and category, scale and nominal variables, respectively.

Individual socio-demographic characteristics included the only variable available that is the role of the user, a nominal variable.

We used the SPSS program to estimate the destination choice model under the two different geographic approaches, which statistical inference, namely at the level accuracy of the model, is guaranteed through the statistical tests included. The model was not identified through tests on the data, but it was previously identified by the researchers for being the most adequate with this kind of variables and when the outputs are of the discrete choice kind.

4 Experimental Results – Interpretation

4.1 Destination Choice Models (Multinomial Logistic Regression)

First Model- Territorial Division Sample

General

The results indicated a good fit of the final model. The model explained 49.0% (Nagelkerke R²) of the variance in the destination choice and correctly classified 82.4% of cases. In addition, the final model is significant at a significance level of $\alpha = 0.050\%$ which means that the full model statistically significantly predicts the dependent variable better than the intercept-only model.

In Table 2, we can observe that three of the six variables presented significance to compose the final model. The variables “distance”, “time spent” and “weekday vs weekend” had a significant contribution ($p < 0.05$), but not “check-ins”, “mode” and “role”.

Trip-Related Characteristics

The model indicates that “distance”, “time spent” and the variable representing the week phase, that is, whether the trip would have occurred during non-routine hours on a weekday or during the weekend, had significance in the choice of destination category type. It was concluded that it was more likely that a user chooses to go to an “Event” than to an “Arts & Entertainment” POI if the trip occurred on a weekday. Additionally, it could be stated that it is 9.959 times more likely that the user chooses to go to an “Event” than to an “Art Exposition/Market” if it was a weekday. Furthermore, the variable “time spent” was shown to be significant. It was found that that increasing the time spent in a trip was associated to a slight likelihood of the user choosing to go visit an “Arts & Entertainment” POI.

Event/POI Attributes

Besides the category, the only variable studied in the model related to the characteristics of the events/POIs consisted of the number of check-ins. This proved not to be significant in the choice of the destination type.

Socio-demographic Characteristics

The only variable related to the individual socio-demographic characteristics introduced in the model was the role. This proved not to be significant.

Second Model-Clusters Division Sample

General

This model only considered trips that culminate in a cluster. In addition, the cluster itself is considered as an independent variable to add to the multinomial logistic regression. This new event/POI attributes related characteristic intuitively represents not only a means to filter trips that are not so important in the sense they are not going to hotspots in the city, but at the same time create a categorical variable which groups trips with close destinations (to the same end cluster). Additionally, we also registered if the trip started in a cluster (origin cluster).

Table 2. Multinomial logistic regression with territorial division.

	Event			Food			Nightlife Spot			Outdoors & Recreation		
	B	Exp(B)	Sig.	B	Exp(B)	Sig.	B	Exp(B)	Sig.	B	Exp(B)	Sig.
<i>Distance</i>	.000	1.000	.005	.000	1.000	.010	.000	1.000	.112	.000	1.00	.212
<i>Time</i>	.000	1.000	.314	.001	1.001	.007	.000	1.000	.527	.000	1.00	.653
<i>Check-ins</i>	.000	1.000	.970	.000	1.000	.811	.000	1.000	.996	.000	1.00	.591
<i>Weekday vs Weekend (ref. Weekday)</i>												
Weekend	2.298	9.959	.000	.177	1.194	.664	.394	1.483	.498	.089	.375	.870
<i>Mode (ref. Sustainable modes)</i>												
Car	0.116	1.123	.785	-.507	602	.229	-.621	.537	.276	.730	2.075	.223
<i>Role (ref. Non-teaching personal)</i>												
Student 1 st cycle	-1.423	.241	.859	-1.188	.305	.889	.773	2.166	.956	-7.102	1.308E-10	.374
Student 2 nd cycle	-2.072	.126	.796	-1.694	.184	.842	-.323	.724	.982	-6.941	1.578E-10	.384
Student 3 rd cycle	.676	1.965	.935	1.259	3.523	.886	.497	1.644	.972	.631	1.862E-07	.939
Researcher	.494	1.638	.972	.172	1.188	.990	3.137	23.027	.866	6.673	3.388E-10	.646
Teacher	.629	1.876	.937	-1.130	.323	.894	.652	1.919	.963	-6.194	3.297E-10	.438

Reference category is “Art Exposition/Market”. Significance level: $p < 0.05$

The results indicated a good fit of the final model. The model explained 28.7% (Nagelkerke R2) of the variance in the destination choice and correctly classified 84.8% of cases. In addition, the final model is significant at a significance level of $\alpha = 0.050\%$ which means that the full model statistically significantly predicts the dependent variable better than the intercept-only model.

In Table 3, differently from the previous experiment, five of the eight variables presented significance to compose the final model. For the simplicity of the table, only when a cluster is considered significant is shown. As the previous model, the variables “distance”, “time spent” and “weekday vs weekend” had a significant contribution ($p < 0.05$), but not “check-ins”, “mode” and “role”. Additionally, the new variables “cluster destination” and “cluster origin” also contribute significantly to the model.

Trip-Related Characteristics

This model indicated the same conclusion for the variables “distance”, “time spent” and “weekday/weekend”, showing that they had significance in the choice of destination category type. Nevertheless, this time, there are more categories where these variables are significant. This model, besides the conclusions made previously, now concludes that it was more likely that a user chooses to go to a “Food” than to an “Arts & Entertainment” POI if the trip occurred on a weekday.

Event/POI Attributes

Besides the category, the variables studied in the model related to the characteristics of the events/POIs consisted of the number of check-ins, cluster origin and cluster destination. These last two new variables proved to be significant in the choice of destination type. It was concluded that it is 1.044 times more likely that the user chooses to go to a “Nightlife Spot” POI than to an “Art Exposition/Market” if the destination of the trip is a density-based cluster.

Socio-demographic Characteristics

The only variable related to the individual socio-demographic characteristics introduced in the model was the user role. This proved not to be significant even in this model.

4.2 Comparing Results

We can see that for small samples, the differences are not substantial. For easier analysis, we put side-by-side the main results obtained through the two approaches, including the significant independent variables (S) and non-significant (N) for both models, as well as the respective percentages of correctly classified cases (Table 4). According to Table 4, both models suggest that the only independent variables that contributed significantly were the day of the week, the travel distance, and the time spent traveling. However, there is a slight improvement in the percentage of cases correctly classified by cluster analysis than by territorial division analysis, and the fact that the new variables (destination and origin of the cluster) are also significant. We believe that this improvement is due to the fact that this analysis is more focused on the POI/event, once the way the data is collected in clusters technique is different than in the territorial division.

Table 3. Multinomial logistic regression with clusters.

	Event			Food			Nightlife Spot			Outdoors & Recreation		
	B	Exp(B)	Sig.	B	Exp(B)	Sig.	B	Exp(B)	Sig.	B	Exp(B)	Sig.
<i>Distance</i>	.000	1.000	.000	.000	1.000	.000	.000	1.000	.002	.000	1.00	.117
<i>Time</i>	.000	1.000	.277	.001	1.001	.015	.000	1.000	.802	.000	1.00	.574
<i>Check-ins</i>	.000	1.000	.988	.000	1.000	.702	.000	1.000	.687	.000	1.00	.503
<i>Weekday vs Weekend (ref. Weekday)</i>												
Weekend	2.535	12.620	.001	-1.132	.322	.049	-.446	.640	.656	.665	1.944	.460
<i>Mode (ref. Sustainable modes)</i>												
Car	-.505	.604	.360	-.989	.372	.076	.219	1.245	.776	.621	1.860	.474
<i>Role (ref. Non-teaching personal)</i>												
Student 1 st cycle	-2.341	.272	.735	-1.420	.242	.862	-.772	.462	.958	.519	1.680	.963
Student 2 nd cycle	-3.687	.096	.593	-2.296	.101	.778	-2.466	.085	.866	-.751	.472	.956
Student 3 rd cycle	-.168	.025	.983	1.387	4.003	.877	-.148	.862	.992	10.749	4.66E+04	.936
Researcher	-.010	.845	1.000	-.130	.878	.994	1.518	4.563	.949	17.556	4.21+E07	.286
Teacher	-.037	.990	.996	-.741	.476	.927	-.545	.580	.970	.939	2.577	.413
<i>Cluster destination</i>	.001	1.001	.899	.025	1.025	.002	.043	1.044	.000	.011	1.011	350
<i>Cluster origin</i>	-.001	.999	.856	-.003	.997	.670	.022	1.022	.012	-.015	.985	.243

This study was a first attempt to extract useful information from georeferenced data from social media to study non-routine mobility. In addition to allowing us to contribute to the knowledge of this type of travel by studying the factors that influence the choice of event/POI category, the results obtained lead us to believe that a further analysis with a larger sample size and not as biased towards the type of people who took the trips would be worthwhile, and we would most likely have even better results.

Table 4. Multinomial logistic regression: comparing territorial with clusters.

	Territorial division (2934 trips)	Clusters (2091 trips)
Role	N	N
Transport mode	N	N
Weekday vs Weekend	S	S
Travel distance	S	S
Time spent	S	S
Check-ins	N	N
Cluster destination	N.A.	S
Cluster origin	N.A.	S
Overall percentage	82.4%	84.8%

5 Conclusion and Future Work

This paper discusses two techniques for extracting information for the study of non-routine travel, namely to characterize their purpose. With the increasing availability of social network and crowd sensed data unsupervised techniques are required to detect the city's main attractive regions in the city during those periods in order to anticipate the demand for transportation. Social events and Points-of-Interest are a good enrichment of traditional studies of these mobility patterns. This work compares the territorial division commonly used in studies of mobility patterns with the automatic recognition of clusters of dense concentrations of POIs or events serving origin-destination region of those trips. In this way, only the trips that are destined to reach a pole of attractiveness are considered in the new approach.

Two models of multinomial logistic regression were constructed from data available in the Porto Metropolitan Area, considering trips on non-routine periods in weeks and on weekends in a period of one month. Some of the trip-related and event/POI variables used showed to be significant to these models, as opposed to a unique demographic variable used due to anonymity of data that proved not to be significant.

With the clustering technique, it was possible to improve the model that only used municipalities as divisions in the study area: both in the significance of the variables and in the final accuracy. However, there is room for improvement as more variables can be added to the model, such as the period of visit at the destination of a trip. This feature is only available after anonymization if the daily session ID created by the crowd sensing mobile application is used. The dataset of trips used was limited to only one month. As a future approach to use the same methodology applied to a bigger dataset, we can test if the silhouette score can handle large data and see other indexes that can be used to identify the best configuration to each clustering algorithm. The choice of the index is important because that can be used to improve the validation time.



References

1. Santos, P.M., et al.: PortoLivingLab: an IoT-based sensing platform for smart cities. *IEEE Internet Things J.* **5**(2), 523–532 (2018)
2. Grigolon, A.B., Kemperman, A.D.A.M., Timmermans, H.J.P.: Mixed multinomial logit model for out-of-home leisure activity choice. *Transp. Res. Rec. J. Transp. Res. Board* **2343**(1), 10–16 (2013)
3. Gkiotsalitis, K., Stathopoulos, A.: Joint leisure travel optimization with user-generated data via perceived utility maximization. *Transp. Res. Part C Emerg. Technol.* **68**, 532–548 (2016)
4. Steed, J.L., Bhat, C.R.: On modeling departure-time choice for home-based social/recreational and shopping trips. *Transp. Res. Rec. J. Transp. Res. Board* **1706**(1), 152–159 (2000)
5. Tarigan, A.K.M., Kitamura, R.: Week-to-week leisure trip frequency and its variability. *Transp. Res. Rec. J. Transp. Res. Board* **2135**(1), 43–51 (2010)
6. Sánchez, O., Isabel, M., González, E.M.: Travel patterns, regarding different activities: work, studies, household responsibilities and leisure. *Transp. Res. Procedia* **3**, 119–128 (2014)
7. Sener, I., Bhat, C., Pendyala, R.: When, where, how long, and with whom are individuals participating in physically active recreational episodes? *Transp. Lett.* **3**(3), 201–217 (2011)
8. Große, J., Olafsson, A.S., Carstensen, T.A., Fertner, C.: Exploring the role of daily ‘modality styles’ and urban structure in holidays and longer weekend trips: Travel behaviour of urban and peri-urban residents in Greater Copenhagen. *J. Transp. Geogr.* **69**, 138–149 (2018)
9. Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J., Ratti, C.: Understanding individual mobility patterns from urban sensing data: a mobile phone trace example. *Transp. Res. Part C Emerg. Technol.* **26**, 301–313 (2013)
10. Anda, C., Erath, A., Fourie, P.J.: Transport modelling in the age of big data. *Int. J. Urban Sci.* **21**(Suppl. 1), 19–42 (2017)
11. Thomas, T., Geurs, K.T., Koolwaaij, J., Bijlsma, M.: Automatic trip detection with the Dutch mobile mobility panel: towards reliable multiple-week trip registration for large samples. *J. Urban Technol.* **25**(2), 143–161 (2018)
12. Sun, Y.: Investigating ‘Locality’ of intra-urban spatial interactions in New York City using foursquare data. *ISPRS Int. J. Geo-Inf.* **5**(4), 43 (2016)
13. Huang, A., Gallegos, L., Lerman, K.: Travel analytics: understanding how destination choice and business clusters are connected based on social media data. *Transp. Res. Part C Emerg. Technol.* **77**, 245–256 (2017)
14. Zhou, Y., Lau, B.P.L., Yuen, C., Tuncer, B., Wilhelm, E.: Understanding urban human mobility through crowdsensed data. *IEEE Commun. Mag.* **56**(11), 52–59 (2018)
15. Zhang, J., Guo, B., Chen, H., Yu, Z., Tian, J., Chin, A.: Public sense: refined urban sensing and public facility management with crowdsourced data. In: 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th International Conference on Autonomic and Trusted Computing and 2015 IEEE 15th International Conference on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), pp. 1407–1412 (2015)
16. Quadri, C., Zignani, M., Gaito, S., Rossi, G.P.: On non-routine places in urban human mobility. In: 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), pp. 584–593 (2018)
17. Sumudu Hasala, M., et al.: Identifying points of interest for elderly in Singapore through mobile crowdsensing. In: Proceedings of the 6th International Conference on Smart Cities and Green ICT Systems, pp. 60–66 (2017)

18. Gan, Z., Yang, M., Feng, T., Timmermans, H.: Understanding urban mobility patterns from a spatiotemporal perspective: daily ridership profiles of metro stations. *Transportation (AMST)* 1–22 (2018)
19. Czepkiewicz, M., Heinonen, J., Ottelin, J.: Why do urbanites travel more than do others? A review of associations between urban form and long-distance leisure travel. *Environ. Res. Lett.* **13**(7), 073001 (2018)
20. Rodrigues, J.G.P., Pereira, J.P., Aguiar, A.: Impact of crowdsourced data quality on travel pattern estimation. In: *Proceedings of the First ACM Workshop on Mobile Crowdsensing Systems and Applications - CrowdSenSys 2017*, pp. 38–43 (2017)
21. Rodrigues, J.G.P., Aguiar, A., Queiros, C.: Opportunistic mobile crowdsensing for gathering mobility information: lessons learned. In: *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1654–1660 (2016)
22. Rendón, E., Abundez, I., Arizmendi, A., Quiroz, E.M.: Internal versus External cluster validation indexes. *Int. J.* **5**(1), 27–34 (2011)



A Flexible and Scalable Architecture for Human-Robot Interaction

Diego Reforgiato Recupero^(✉), Danilo Dessì, and Emanuele Concas

Department of Mathematics and Computer Science,
University of Cagliari, Cagliari, Italy

{diego.reforgiato,danilo.dessi}@unica.it, e.concas11@studenti.unica.it

Abstract. Recent developments and advancements in several areas of Computer Science such as Semantic Web, Natural Language Understanding, Knowledge Representation, and more in general Artificial Intelligence have enabled to develop automatic and smart systems able to address various challenges and tasks. In this paper, we present a scalable and flexible humanoid robot architecture which employs artificial intelligent technologies and developed on top of the programmable humanoid robot called Zora. The framework is composed by three different modules which enable the interaction between Zora and a human for tasks such as Sentiment Understanding, Question-Answering, and automatic Object Recognition. The framework is flexible and extensible, and can be augmented by other modules. Moreover, the embedded modules we present are general, in the sense that they can be easily enriched by adding training resources for the presented sub-components. The design of each module consists of two components (i) a front-end system which is responsible for the interaction with humans, and (ii) a back-end component which resides on server side and performs the heavy computation.

Keywords: Human-robot interaction · Natural Language Understanding · Semantic Web · Sentiment analysis · Artificial intelligence · Zora

1 Introduction

Nowadays we are assisting to the spread of robots in many fields to perform tasks in the place of humans. For example they may be found for performing hazardous tasks (e.g. the exploration of volcanoes), helping people with pathological health problems [1] and executing a fairly wide range of tasks (e.g., house cleanings, painting, packaging) with very little human intervention [8]. Next robot generations fit into every day human life raising the need to build systems for a simple engagement between robots and humans. Therefore, robots should be able to accurately assess people's actions and requests, and perform tasks accordingly.

In a human-robot interaction, people typically have high expectations about robot abilities in understanding their requests, contents and feelings [13]. For

building this kind of robots, novel platforms that give smart and friendly behaviors to robots must be developed so that robots can have a human appearance.

In the past, robots social behaviour has been simulated through remote control systems so that humans that interact with the robots had the best possible interaction. This approach clearly does not work at large scale because it needs manual intervention of humans [3], limiting the spread of this technology in our daily life. Hence, the study of systems to provide autonomous social robots that can directly interact with people has taken hold for providing services in common tasks day by day.

Recent developments and advancements in several areas of Computer Science such as Semantic Web, Natural Language Understanding, and Knowledge Representation, and more in general in the Artificial Intelligence field, have enabled the development of automatic and smart systems able to address various challenges and tasks, providing robots with much better autonomous and social behaviours. Therefore, in this paper we

- introduce an extensible framework which is current composed by three modules for enabling Zora, the humanoid robotic platform we have employed, to have different skills;
- we provide Zora with a Sentiment Understanding engine for capturing the feelings of a person who is interacting with the robot;
- we provide Zora with a Question-Answering engine to engage the robot in a clever dialog with a person;
- we provide Zora with an Object Recognition engine to detect objects.

Each module exploits results of different research problems [2] we have previously addressed through state-of-the-art Deep Learning technologies on various datasets. In our approach we included the best models and resources we were able to obtain, making a first step for a smart framework for Zora.

2 Background

Machine Learning methods have proved their validity in many fields, and particularly interesting developments have been carried out in Natural Language Understanding and Object Detection. The Natural Language Understanding benefit of the use of Semantic Web technologies such as Stanford CoreNLP [11], Framester [6], FRED [7], and Word Embeddings [12] to move the data representation from simple text to concepts, enabling better models to capture the meaning of natural language expressions. Together with recent developments in Deep Learning they all have obtained great results in those tasks that require human-based cognitive abilities. The powerful technologies may be exploited for powering Conversational Agents, developing systems that can interpret and respond to the input of a user in a smart way. Furthermore, Object Recognition has been also improved with Deep Learning based approaches which are nowadays able to build models that can recognize object with high level of accuracy. The idea behind this paper took inspiration from [9] where a framework able

to perform semantic interpretation with the NAO robot was proposed. In the study, authors focused mainly on ontologies both for mapping the user speech and to execute actions accordingly. Differently, we have based our framework on Deep Learning based modules and Semantic Web technologies for providing smart services through the robot Zora.

3 The Proposed Framework

In this section we describe our framework composed by three modules. Each module is composed by a Choregraphe Component (CC) and a Server Side Support Component (SSSC). CC is very light in terms of required resources (physical space and memory demand), and is directly uploaded and executed by Zora. To note that Zora is a robot which is based on the same infrastructure of NAO, a humanoid robot designed by the Aldebaran Robotics company in France and that presents an intuitive user interface that can be equipped with software packages providing skills to the robot. The SSSC is responsible for the heavy computation on a dedicated server. It runs AI, NLP, Semantic Web and Computer Vision, and any other kind of applications that may be necessary to develop new modules. Each module runs independently from the others, and more than one module at time can be included in the framework. However, a priority order of the modules is set in the framework so that when a user input is received only one module is executed. Furthermore, each module can be activated by the user through the use of *command-words* so that he/she can decide

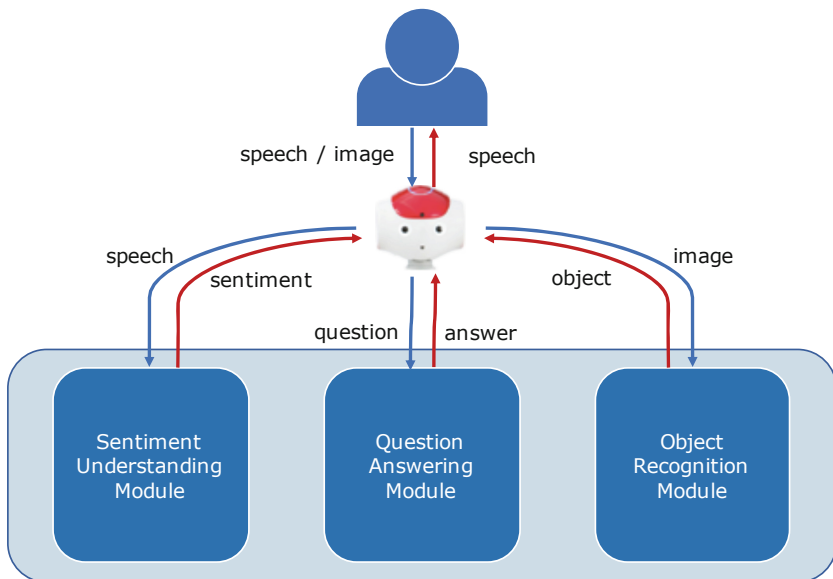


Fig. 1. The schema of the interaction model using Zora.

when to start the interaction with the robot which leverages one of the uploaded modules. The current architecture is public accessible¹ (Fig. 1).

3.1 Sentiment Understanding Module

The Sentiment Understanding [5] Module gives Zora the ability to classify a natural language user’s input. It first converts the spoken input of the user in simple text through a speech-to-text internal tool, and then sends the text to the SSSC. The SSSC performs a sentiment polarity detection task adopting a Deep Learning model which classifies the input in one of the following polarity classes: *positive*, *negative*, and *neutral*. A schema of the Deep Learning model is shown in Fig. 2(a). The textual data is represented by means of word embeddings which are subsequently fed into the Deep Learning model. The BiLSTM are levels that implement Bidirectional Long-Short Term Memory neural networks that capture patterns of data in both forward and backward direction and, at the same time, can consider long sequences of data by relating the first part of a sentence with its last part. Finally, the model presents an Attention layer which allows to refer to data previously processed instead than forcing the input of the final layer to a single vector. The last layer, *Dense*, combines the result of the whole computation and returns a single output. Afterwards, the result is sent back to the CC which delivers the predicted sentiment to the user. The sentiment model has been trained using the *The Large Movie Review Dataset* [10].

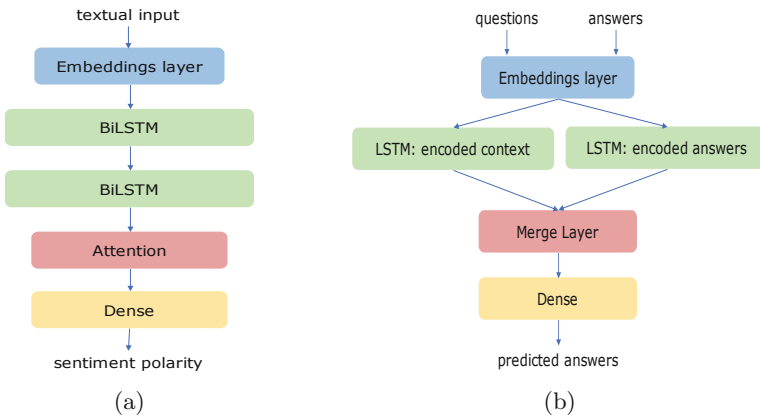


Fig. 2. Deep Learning models of (a) Sentiment Understanding (b) Question-Answering.

3.2 Question-Answering Module

With this module, Zora is able to reply to the user with meaningful responses based on the dialog history with several other users used as training set. The CC

¹ <https://github.com/hri-unica>.

interacts with the user through Zora microphones and sends the transcription of his/her voice to the SSSC. The SSSC takes the text, and uses it as an input of a model that has been previously trained. The model is a Deep Learning hierarchy that has been previously trained on the *Cornell Movie Dialogs Corpus* [4] representing the data with Word Embeddings. A schema of the adopted Deep Learning hierarchy is shown in Fig. 2(b). It captures syntactical and semantic patterns from a sequence to sequence input defining a context. It uses a LSTM layer to capture the semantic of questions and answers, a Merge layer to bind possible answers to questions, and a Dense layer to predict the best token for the answer. As for the Sentiment Understanding Module, the textual input is represented by word embeddings. The response predicted by the SSSC is sent to the CC which delivers the message to the user.

3.3 Object Recognition Module

For this module, the Zora front camera is used to take a picture of size 640×480 . The picture is sent to the SSSC which implements an internal TensorFlow model that is currently able to recognize 91 types of general objects (such as *train*, *person*, *animal*). For the categories *cat* and *dog*, it also recognizes a more fine-grained class (e.g. a dogs breed like *american bulldog*). The SSSC uses two classifiers. One classifier is used to predict the general class. In case the predicted class is one between *dog* or *cat*, it uses a second classifier to further recognize the breed. The general object recognition model has been trained with the Microsoft COCO dataset², while the Oxford-IIIT-Pet dataset³ has been employed for the breed recognition model training. When the elaboration of data has ended by the SSSC and the result has been sent to the CC, the CC tells the user which objects have been recognized. If no objects are recognized, the robot asks the user which objects have been shown and user's annotations are sent to the SSSC which stores the new data. When the interaction is interrupted by the user, the SSSC automatically performs a new training of the model integrating the updated annotated sets with the new data.

4 Conclusions and Future Work

In this paper we have introduced a smart framework which leverages our research in various domains to equip Zora, a completely programmable and autonomous humanoid robot built on top of NAO, with smart human-based skills. We have described three modules which exploit Deep Learning models and Semantic Web technologies to manipulate the input of a user enabling Zora to respond accordingly. The modules allow Zora performing sentiment analysis, generating automatic answers to user's natural language query as in a real dialog, and detecting objects. All modules are based on Deep Learning technologies which are executed on server-side so that the robot is free from the heavy computation. Summing

² <http://cocodataset.org>.

³ <https://www.robots.ox.ac.uk/~vgg/data/pets/>.

up, each module has (i) a CC which runs on Zora allowing the interaction with the user, and manages inputs and outputs, and (ii) a SSSC which trains Deep Learning model and performs a prediction task on the user's input. As future works, we mainly aim at integrating the framework with modules which allow Zora performing body actions based on the input of the user and interacting with home automation devices such as Google Home. Moreover, we plan to perform different tests to prove the effectiveness and usefulness of the proposed framework.

References

1. Asprino, L., Gangemi, A., Nuzzolese, A.G., Presutti, V., Recupero, D.R., Russo, A.: Ontology-based knowledge management for comprehensive geriatric assessment and reminiscence therapy on social robots. *Data Science for Healthcare*, pp. 173–193. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-05249-2_6
2. Atzeni, M., Reforgiato Recupero, D.: Deep learning and sentiment analysis for human-robot interaction. In: Gangemi, A., Gentile, A.L., Nuzzolese, A.G., Rudolph, S., Maleshkova, M., Paulheim, H., Pan, J.Z., Alam, M. (eds.) *ESWC 2018*. LNCS, vol. 11155, pp. 14–18. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-98192-5_3
3. Breazeal, C., Takanishi, A., Kobayashi, T.: Social robots that interact with people. In: Siciliano, B., Khatib, O. (eds.) *Springer Handbook of Robotics*. Springer, Berlin, Heidelberg (2008). https://doi.org/10.1007/978-3-540-30301-5_59
4. Danescu-Niculescu-Mizil, C., Lee, L.: Chameleons in imagined conversations: a new approach to understanding coordination of linguistic style in dialogs. In: *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*, pp. 76–87 (2011)
5. Dridi, A., Reforgiato Recupero, D.: Leveraging semantics for sentiment polarity detection in social media. *Int. J. Mach. Learn. Cybern.* **10**(8), 2045–2055 (2019). <https://doi.org/10.1007/s13042-017-0727-z>. cited By 3
6. Gangemi, A., Alam, M., Asprino, L., Presutti, V., Recupero, D.R.: Framester: a wide coverage linguistic linked data hub. In: Blomqvist, E., Ciancarini, P., Poggi, F., Vitali, F. (eds.) *EKAW 2016*. LNCS (LNAI), vol. 10024, pp. 239–254. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49004-5_16
7. Gangemi, A., Presutti, V., Reforgiato Recupero, D., Nuzzolese, A.G., Draicchio, F., Mongiovì, M.: Semantic web machine reading with fred. *Seman. Web* **8**(6), 873–893 (2017)
8. Graetz, G., Michaels, G.: Robots at work. *Rev. Econ. Stat.* **100**(5), 753–768 (2018)
9. Kobayashi, S., Tamagawa, S., Morita, T., Yamaguchi, T.: Intelligent humanoid robot with Japanese wikipedia ontology and robot action ontology. In: *Proceedings of the 6th International Conference on Human-Robot Interaction*, pp. 417–424 (2011)
10. Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, C.: Learning word vectors for sentiment analysis. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 142–150 (2011)

11. Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., McClosky, D.: The stanford corenlp natural language processing toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55–60 (2014)
12. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint [arXiv:1301.3781](https://arxiv.org/abs/1301.3781) (2013)
13. Piccolo, L., Mensio, M., Alani, H.: Chasing the chatbots: Directions for interaction and design research (2019)



Toward Supporting Food Journaling Using Air Quality Data Mining and a Social Robot

Federica Gerina, Barbara Pes, Diego Reforgiato Recupero,
and Daniele Riboni^(✉)

Department of Mathematics and Computer Science, University of Cagliari,
Cagliari, Italy
gerinafederica@gmail.com, {pes,diego.reforgiato,riboni}@unica.it

Abstract. Unhealthy diet is a leading cause of health issues. A powerful means for monitoring and improving nutrition is keeping a food diary. Unfortunately, frail people such as the elderly have a hard time filling food diaries on a continuous basis due to forgetfulness or physical issues. For this reason, in this paper we investigate the integration of nutrition monitoring in a robotic platform. A machine learning module detects cooking activities based on air quality sensor data. When cooking is detected, a social robot interacts with the user to fill the food diary through a conversational interface. We report our experience on the development of a partial prototype of our system. Moreover, we illustrate the results of preliminary experiments with annotated sensor data gathered over one month from a real-world apartment.

Keywords: Healthcare · Context-aware computing · Social robots

1 Introduction

The 2018 Global Nutrition Report¹ of the World Health Organization reveals that malnutrition determines more health issues than any other cause. Diet data analysis is of foremost importance to evaluate the healthiness on an individual's nutrition and for setting up interventions. A powerful tool for acquiring diet data and augmenting self-awareness is *food journaling*, which consists of filling a diary of eaten food and its quantity at each meal [6]. Of course, manually keeping a food diary in the long term is tedious, since diary annotation interferes with the current activity. In order to assist the individual in filling a food diary, different solutions have been proposed, which exploit mobile apps and smartphone sensors [1, 5, 8, 11]. However, mobile apps for food journaling still require considerable user's effort, and are often unsuitable for frail people, including the elderly. On the other hand, assistive robots are increasingly used as support for frail people. Plenty of research has been carried out on assistive robots [3]. Whereas,

¹ <https://www.who.int/nutrition/globalnutritionreport/en/>.

in the past, assistive robots mainly helped people through physical interaction, in recent years, the social interaction with the user has become the emerging direction to increase the levels of engagement. As such, researchers have widely investigated user engagement with a robot companion and found out that it is much higher than with respect to any other device or digital assistant [2, 10].

In this paper, we tackle the challenging issue of integrating a system for food journaling in a social robot platform addressed to frail people. We propose a novel architecture for acquiring food diary data through human-robot interaction, exploiting conversational interfaces, advanced air quality sensors, and artificial intelligence methods. Our system relies on data captured by an air quality monitor to recognize food preparation using machine learning. Compared to other cooking recognition systems based on cameras, our data acquisition system is unobtrusive and privacy-conscious. Whenever the supervised classifier predicts a cooking activity, our system informs a social robot in the home, which interacts with the user to fill the food diary through a conversational interface.

The complete realization of the system is challenging. Currently, we have developed a partial prototype of our system. For the sake of this work, we concentrate on the initial design of the platform, on the acquisition and processing of sensor data, and on machine learning for recognizing the cooking activity that triggers robot-human interaction. Advanced solutions for conversational interfaces, usability, and diet analysis, including calorie count and adaptive interfaces for supporting behavior change, will be addressed in future work.

The paper is structured as follows. We illustrate our architecture and prototype in Sect. 2. Acquisition of air quality sensor data and feature engineering are covered within Sect. 3. Preliminary experimental results are reported in Sect. 4, whereas Sect. 5 concludes the paper.

2 Preliminary Architecture and Prototype

Figure 1 shows the preliminary architecture of our system. A machine learning module contains the trained model. That module exposes REST APIs to classify as *cooking* or *non cooking* a new collected record of sensor data acquired from an air quality monitor in the kitchen. A software agent periodically collects sensor data and calls the REST APIs of the machine learning module. If the new read data is classified as *cooking*, the robot starts interacting with the user asking what food he/she is preparing. Once the user replies, the robot performs speech to text processing and sends the extracted food information to the food journaling database. Note that, in the current implementation of our system, the speech interaction to acquire food journaling data is over-simplistic, being based on a simple question-answer paradigm. Since most food journaling applications require detailed information about the kind and quantity of food, we will investigate a more sophisticated conversational agent for food journaling in future work. Voice-based identification methods will also be used to recognize the inhabitant, in case of multi-resident homes. Moreover, we will use active learning methods to improve the trained model based on new data and user's feedback.

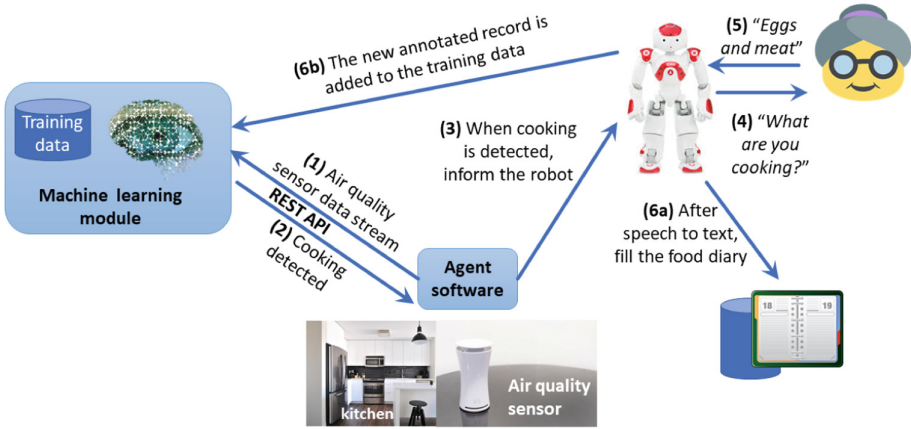


Fig. 1. Overall architecture of our system.

We have developed a preliminary prototype of the system in order to evaluate its feasibility. We have used a uHoo device² to continuously acquire air quality data, including temperature, humidity, carbon dioxide, volatile organic compounds, particulate matter. Figure 2a shows a uHoo device used in our experimental setup. The sensor has a detection frequency of one reading per minute. It provides a wireless network interface and electrical connection to avoid battery exhaustion, and open APIs for acquiring the sensor data in real time. For the social robot platform, we adopted a Zora robot³, which uses the same robotic infrastructure of Nao, but adds an extremely simple and intuitive user interface. Figure 2b shows an image of Zora. In order to capture the voice of the user when he/she speaks, the robot is equipped with four microphones. The robot can therefore record the human voice, which is contextually analyzed and transformed into text by a speech recognition module. We are currently relying on cloud computing systems for speech recognition in order to improve the accuracy of the speech-to-text process. In fact, this allows us pre-processing the sound recorded by Zora and removing noise (e.g background noise, fan noise, etc.), which may compromise the result. As such, the resulting audio file is sent to IBM Watson Speech to Text to perform speech recognition.

3 Acquisition and Processing of Air Quality Sensor Data

In this section, we explain how we acquire and process air quality sensor data in order to recognize the preparation of food. The indoor air quality monitor deployed in the kitchen is in charge of providing a stream of real-time sensor data to the MACHINE LEARNING module. That module performs FEATURE EXTRACTION, to build feature vectors based on statistics computed on the streaming

² <https://uhooair.com/>.

³ <https://www.youtube.com/watch?v=IO52sLF-u.4&t=1s>.

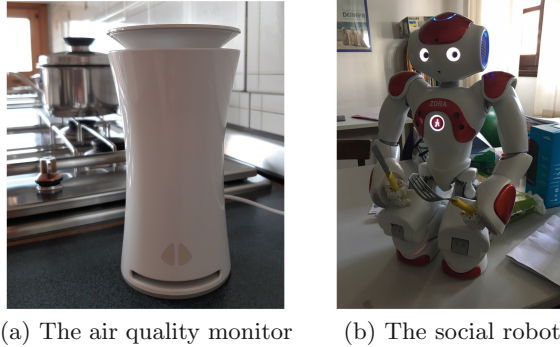


Fig. 2. The devices used in our system.

data. The feature vector is passed to the ONLINE RECOGNITION module, which uses a machine learning classifier to detect whether the user is cooking or not. The classifier is trained in advance using a labeled training set of sensor data acquired during cooking and non-cooking activities. Finally, the prediction is communicated to the robot.

From the analysis of the daily trends of air quality data, it emerges that, in order to distinguish cooking from non-cooking activities, it is important to analyse the trend of values, not only the absolute values. For this reason, we engineered features taking into account not only the absolute values or averages, but also the difference between the current value and the preceding/succeeding values. In particular, we build features considering the differences between the most recent value and the one in the previous/next 5, 10, 15, 20, and 25 min. Obviously, the use of features computed considering succeeding values determines a delay in the recognition process, but this delay is acceptable for our application. We also use statistical features considering the average, minimum, and maximum value in the last 5 min, as well as the standard deviation of those values. Finally, we also compute the current time of the day as the number of minutes that passed from midnight. We use 114 features in total.

4 Experimental Evaluation

In this section, we report the results of preliminary experiments carried out with real-world data. Our dataset is published on the Web⁴. It is composed of 55768 data readings taken at each minute during more than one month in total from one volunteer in his apartment. Each data record is labeled with the current activity, that can take two values: 1 if the user is cooking a meal (positive class), 0 otherwise (negative class). The number of records with activity set to 1 is 3451, while the remaining 52317 are set to 0; hence, classes are strongly unbalanced.

⁴ <http://sites.unica.it/domusafe/single-apartment/>.

Table 1. Experimental results using the Random forest classifier.

Precision	Recall	Specificity	F_1 score	G-mean	Cooking	Notcooking	← Classified as
0.759	0.748	0.985	0.754	0.858	2529	852	Cooking
					802	51036	Notcooking

(a) Evaluation metrics

(b) Confusion matrix

We have used the Weka toolkit [7] and experimented with different classifiers applying 10 fold cross-validation without shuffling. For lack of space, we report only the results achieved by the most effective classifier in our pool, which was Random forest [4]. The metrics used to evaluate the model are the standard ones of precision, recall, specificity, F_1 score (harmonic mean of precision and recall), and G-mean (geometric mean of specificity and recall) [9]. Table 1a summarizes the results. Our system recognized the positive class (i.e., Cooking activities) with more than 75% precision and almost 75% recall, achieving an F_1 score of 0.754. On the other hand, more than 98% of the negative instances (i.e., Not cooking activities) were correctly classified, as shown by the specificity value. Globally, the trade-off between true positive and true negative rates, as expressed by the G-mean, is 0.858. Results can be inspected in more detail in the confusion matrix reported in Table 1b.

Overall, we consider the achieved classification performance satisfactory. Indeed, we are interested in recognizing instances of cooking that may last for several minutes; hence, a certain rate of minute-by-minute misclassifications may be acceptable. However, experiments have been conducted with a single user dataset. With training and test data acquired from different individuals in different environments, we expect lower recognition rates. Indeed, the age and number of inhabitants has an impact on the kind and quantity of cooked food, and consequently on the change in air quality conditions determined by cooking. The topology of the home also has an impact on air quality data. For instance, if inhabitants consume the meal within the kitchen, their presence determines an increase of temperature and CO_2 levels even after cooking has ended. On the other hand, if the inhabitants consume the meal in a different room, the CO_2 and temperature levels in the kitchen decrease as soon as the cooking activity is finished. This variability may have a negative impact on recognition rates.

5 Conclusion and Future Work

In this paper, we have introduced a novel system to support food journaling, addressed to frail people living alone. Our system relies on advanced air quality sensors for cooking recognition, and on a social robot for interactively filling the food diary when cooking is recognized. We have developed a partial prototype using a social robot and an air quality monitor, and performed preliminary experiments to assess the effectiveness of our cooking recognition method.

Our preliminary work shows several challenges we had to face and many more that we still need to address. First of all, we will investigate methods to improve the classification performance of our cooking recognition system. Since both the topology of the home and the characteristics of inhabitants (including their number and age distribution) affect the air quality conditions at cooking time, we will investigate and experiment techniques to couple our data-driven method with a knowledge-based one, to fine-tune recognition to home's and inhabitants' characteristics. We will also investigate the integration of other sensors attached to furniture to increase recognition rates. Since not all meals are taken at home, the system could be extended with mobile apps for food journaling. Future work also includes the definition of an effective and engaging conversational interface for interactively filling the food diary.

Acknowledgements. This research was partially funded by the EU's Marie Curie training network PhilHumans - Personal Health Interfaces Leveraging HUMAN-MACHINE Natural interactionS (grant number 812882).

References

1. Amft, O., Stäger, M., Lukowicz, P., Tröster, G.: Analysis of chewing sounds for dietary monitoring. In: Beigl, M., Intille, S., Rekimoto, J., Tokuda, H. (eds.) *UbiComp 2005*. LNCS, vol. 3660, pp. 56–72. Springer, Heidelberg (2005). https://doi.org/10.1007/11551201_4
2. Anzalone, S.M., Boucenna, S., Ivaldi, S., Chetouani, M.: Evaluating the engagement with social robots. *Int. J. Soc. Robot* **7**(4), 465–478 (2015)
3. Bemelmans, R., Gelderblom, G.J., Jonker, P., Witte, L.: Socially assistive robots in elderly care: a systematic review into effects and effectiveness. *J. Am. Med. Dir. Assoc.* **13**(2), 114–120.e1 (2010)
4. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001)
5. Cordeiro, F., Bales, E., Cherry, E., Fogarty, J.: Rethinking the mobile food journal: exploring opportunities for lightweight photo-based capture. In: *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pp. 3207–3216. ACM (2015)
6. DiFilippo, K.N., Huang, W.H., Andrade, J.E., Chapman-Novakofski, K.M.: The use of mobile apps to improve nutrition outcomes: a systematic literature review. *J. Telemed. Telecare* **21**(5), 243–253 (2015)
7. Frank, E., et al.: Weka—a machine learning workbench for data mining. In: Maimon, O., Rokach, L. (eds.) *Data Mining and Knowledge Discovery Handbook*. Springer, Boston (2009)
8. Mankoff, J., Hsieh, G., Hung, H.C., Lee, S., Nitao, E.: Using low-cost sensing to support nutritional awareness. In: Borriello, G., Holmquist, L.E. (eds.) *UbiComp 2002*. LNCS, vol. 2498, pp. 371–378. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45809-3_29
9. Norvig, P., Russell, S.: *Artificial Intelligence A Modern Approach*. Prentice Hall Series in Artificial Intelligence. Prentice Hall, Upper Saddle River (2003)
10. Vaufreydaz, D., Johal, W., Combe, C.: Starting engagement detection towards a companion robot using multimodal features. *Robot Auton. Syst.* **75**, 4–16 (2016)
11. Zhu, F., et al.: The use of mobile devices in aiding dietary assessment and evaluation. *J. Sel. Topics Signal Process* **4**(4), 756–766 (2010)



Viewing Experience of Augmented Reality Objects as Ambient Media - A Comparison of Multimedia Devices

Ilhan Aslan^(✉), Chi Tai Dang, Björn Petrak, Michael Dietz, Michael Filipenko, and Elisabeth André

Human-Centered Multimedia Lab, Augsburg University, Augsburg, Germany
{[aslan](mailto:aslan@hcm-lab.de),[dang](mailto:dang@hcm-lab.de),[petrak](mailto:petrak@hcm-lab.de),[dietz](mailto:dietz@hcm-lab.de),[filipenko](mailto:filipenko@hcm-lab.de),[andre](mailto:andre@hcm-lab.de)}@hcm-lab.de

Abstract. Augmented reality objects as ambient media (i.e., media embedded in everyday contexts) are being deployed in diverse contexts, such as home, mobile, and work. Because media experience is mediated by the medium, the viewing experience of augmented reality objects should depend on the type of viewing device. In this paper, we report on a user study with 18 participants and three state-of-the-art devices (i.e., HoloLens, iPad Pro tablet, and iPhone X smartphone), comparing the influence of these mobile viewing devices on feelings of presence and workload during ambient media consumption. Study results include a significant main effect of device on participants ratings of *Possibility to examine* an AR object, which demonstrates that ambient media consumption experiences are determined by the choice of viewing device.

Keywords: Augmented reality · User experience · Multimodality

1 Introduction

There is an ongoing hype in augmented reality (AR) research and near future visions depicted in contemporary movies, which suggest that AR objects will increasingly become part of our everyday lives and serve as ambient media. To compare how different mobile devices impact users' viewing experience of AR Objects we have conducted an empirical study with 18 participants. Our assumption was that different classes of mobile devices (i.e., head-mounted, tablet, smartphone) may influence viewers' feelings of presence and workload when spatio-visually exploring an AR object. To this end, we report and discuss participants' self-reports on feelings of workload and presence, including their dimensions *Possibility to examine* and *Quality of interface*, which show that today's mobile devices seem to provide a different AR object viewing experience in a simple viewing task than the HoloLens as a state-of-the-art head-mounted display. Overall our research demonstrates that the choice of AR Object viewing device significantly matters considering the viewing experiences, and moreover our research describes specific differences in today's state of the art viewing devices with a focus on feelings of presence and workload.



Fig. 1. Images of the treehouse AR object and sketches of the viewing conditions employed to explore the effect of viewing medium on feelings of presence and workload.

2 User Study

The overall goal of the user study was to explore differences in feelings of presence and workload associated with different classes of mobile devices used as mediums/modalities to view AR objects. We targeted three types of state of the art mobile devices: (i) smartphone (ii) tablet, and (iii) head-mounted mixed reality glass, with smartphone and tablet being video-see-through devices in different sizes and the head-mounted device (i.e., HoloLens) being a binocular optical-see-through device (see Fig. 1). Consequently, our hypothesis was that the type of viewing device effects feelings of presence and workload during an AR object viewing task. To be gender inclusive and because some related work suggests potential gender differences in spatio-cognitive abilities (e.g., [1, 2, 4]) we choose to gender-balance our study.

2.1 Participants, Apparatus and Procedure

We recruited 18 participants (aged between 18–40, 9f and 9m) at the university campus. All participants have reported to have no color vision deficiency and were “tech-savy” and accustomed to using mobile devices. Ten of these participants reported to already have varying levels of experience with augmented reality. The study was conducted at an open space inside the university building. The study space was about 40 square meters and away from potential by-passers. The three devices, which we utilized in the user study as viewing mediums were (i) an iPhone X smartphone, (ii) a first generation iPad Pro tablet, and (iii) a first generation Microsoft’s HoloLens. We chose these devices because we believe that they are archetypical and therefore results are likely to be ecologically valid.

For each of the devices an application was developed, which calculates the exact position to place the AR object based on augmented reality software, and thus allows users to visually perceive the AR object as an overlay to the real world (image). The applications also enable users to implicitly (and explicitly) interact with AR objects by moving around the AR object or holding the viewing medium differently. For the iOS devices (i.e., the iPad and iPhone) we used Apple’s ARKit¹ and its plane detection to find a surface, allowing to place the AR

¹ <https://developer.apple.com/arkit/>.

object at the desired location. The application for the HoloLens was developed using Unity² and Microsoft’s HoloToolkit.

For the purpose of self-reporting feelings of presence, we used the Presence Questionnaire (PQ), which is adapted from Witmer and Singer [5] and revised by the UQO Cyberpsychology Lab (2004). The PQ measures overall presence and consists of five sub-scales: realism, possibility to act, quality of interface, possibility to examine, and self-evaluation of performance, which combined measure overall presence. In addition, we used the NASA-TLX questionnaire [3] to explore how the viewing medium may change feelings of perceived workload (i.e., mental workload, physical workload, temporal workload, performance, effort, frustration, and overall workload). A semi-structured interview was also conducted at the end with each participant.

The user study started with welcoming each participant and handing them a written short description of the study procedure, which stated that participants would receive three augmented reality devices, one after another, with which they would be able to observe an AR object positioned at the center of the study space. Participants were informed that their task was to view the AR object for 3 min and try to remember as many details of the AR object as possible since we would ask them a question after each session about a detail of the AR object. At the end of each session participants were asked to provide self-reports on the viewing experience associated with the specific device based on the PQ and NASA-TLX questionnaire. We used counterbalancing (levels of device and gender) to ensure validity of our study. To ensure participants would view the AR object in each of the three repetitions carefully, they were informed that they would be asked a different detail of the AR object after each session. We asked for example “Which color did the door knob have?”. Each participant completed the study in about 45 min.

2.2 Results

In terms of viewing device preferences, 9 (2f, 7m) participants preferred the tablet, 5 (5f) participants preferred the HoloLens and 4 (2f, 2m) preferred the smartphone. The frequency plots in Fig. 2 provide a descriptive overview of the collected data, considering feelings of workload and presence. Considering the mean ratings for presence, the tablet has received the highest mean scores on all dimensions. An exception are ratings of female participants for the dimension *Realism*, who have provided very similar mean ratings for the HoloLens and the tablet. Besides this exception the HoloLens device has received the lowest mean scores in our sample data. Participants seem to have associated low workload with all devices and the viewing tasks. The *Performance* dimension of the NASA-TLX questionnaire seems to be an exception. Participants provided higher ratings for the *Performance* dimension, which measures how much “performance demand” participants felt when using a specific device to complete the viewing task.

² <https://unity3d.com>.

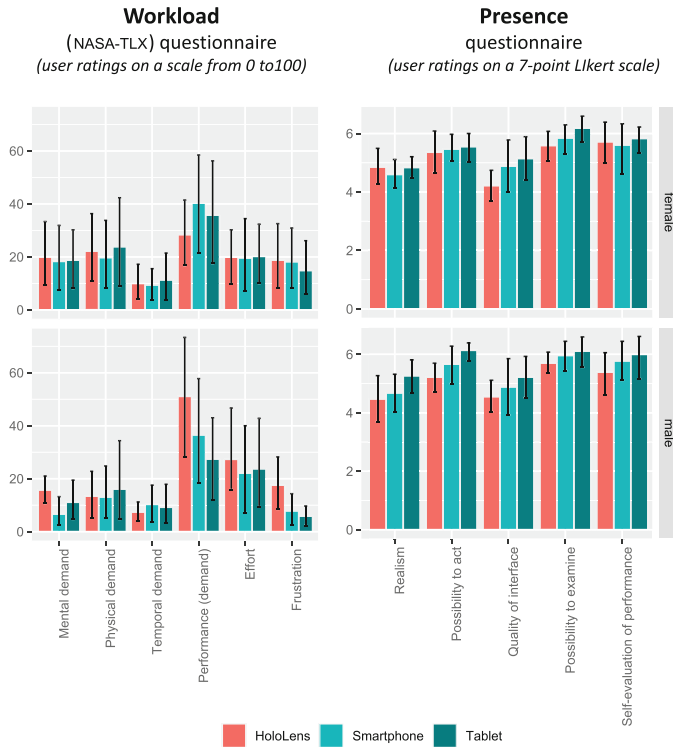


Fig. 2. Overview of the data collected with the workload questionnaire and the presence questionnaire. Descriptive statistics (i.e., mean values and confidence intervals) are presented for all ratings for sub-scales of presence and workload separated by the levels of the two independent variables device and gender. Error bars denote 95% CI.

Overall, the plots in Fig. 2 show that the HoloLens received worse mean scores compared to the other viewing devices from males for both feelings of workload and presence, which could be part of the reason why none of the male participants preferred the HoloLens overall for viewing AR objects.

Statistical Analysis. Table 1 depicts the results of the analysis for which we used ARTool [6]. In order to save space in the table we present only test results, which resulted in p-values below 0.1. While there are some measurements that resulted in a p-value below the 0.1 threshold, including a potential main effect of device on overall presence ($F = 2.63$, $p = 0.087$) only two measurements are statistically significant (i.e., have p-values below 0.05). We found a significant main effect of device on *possibility to exam* ($F = 3.86$, $p = 0.031$) and a significant interaction between gender and device considering *performance* (demand). Post-hoc cross-factor comparisons highlight that this interaction is due to significant differences between HoloLens and tablet and HoloLens and smartphone. If we

Table 1. Overview of results of statistical tests (based on the ARTool [6]) over all participants, considering the overall effect of within-subjects variable *Device* (i.e., HoloLens, Smartphone, and Tablet), between-subjects variable *Gender* (i.e., male and female), and *Interactions of Device and Gender* on measures for perceived presence and workload. Abbreviations used for reporting pairwise comparison: Tablet-HoloLens (T-H), Tablet-Smartphone (T-S), and Smartphone-HoloLens (S-H). F-M is an abbreviation for Female-Male.

Effect of:	Effect on:	F and p	Post-hoc comp.
Device	<i>Overall presence (perceived presence)</i>	F = 2.63 p = .087	
Device	<i>Quality of interface (perceived presence)</i>	F = 3.07 p = .060	
Device	<i>Possibility to examine (perceived presence)</i>	F = 3.86 p = .031 *	
Device	<i>Mental demand (perceived workload)</i>	F = 2.68 p = .083	
Device * Gender	<i>Performance demand (perceived workload)</i>	F = 3.67 p = .036 *	T-H:F-M (p = .03 *); S-H:F-M (p = .08)
Device	<i>Effort (perceived workload)</i>	F = 2.52 p = .095	

try to put the statistics in words, it would mean that the average difference in “performance demand” perceived by male participants between the HoloLens and the other two viewing mediums was significantly opposite to female participants’s perception of the same conditions.

Analysis of Qualitative Data. In the following, we list results of observations of participants’ behaviors during the viewing task and results of the semi-structured interviews which were conducted at the end of each study with each participant. **HoloLens:** When using the HoloLens, participants tended to keep a larger distance to the AR object, which resulted in participants moving at the edge of the study space mostly around the AR object and occasionally moving towards and away from the AR object. Arguments of participants who preferred the HoloLens were focused on reasons why the HoloLens provided a *more natural* and *realistic experience*. Participants argued, for example that “*it felt more natural*” and that it is “*most realistic because one does not have to look through a screen*”. Furthermore, participants stated “*more realistic especially when one has their hands free*” and more realistic because “*the AR object moves [in synchrony] with head movements*”.

Smartphone and Tablet: When using the smartphone or the tablet participants not only moved around the AR object but additionally used the possibility to rotate or move the viewing medium without changing their own position.

Most of the time participants used the handhelds in portrait format, while in the HoloLens' orientation is landscape and fix. Furthermore, handhelds allowed participants to view the AR object from above, below and from the side without always having to move around and instead adjusting their screens position by moving the screen closer, away, to the side, etc. The smartphone was usually held with one hand only and the tablet was always held with both hands. Participants who preferred the tablet focused on benefits of the *larger screen*, arguing “*the viewing angle of the HoloLens and the smartphone is too small*”, “*most details visible*” and “*HoloLens is too demanding and has a too small view, and the smartphone is too small*”. Participants who preferred the smartphone focused on its “*better overall usability*”, arguing that “*it is easiest to change the viewing angle*” or how it is “*better to handle, not as heavy as the tablet*”, and while it is “*small but enough is visible*”.

3 Conclusion

We have reported on an empirical study with users, exploring the influence of state-of-the-art (and off-the-shelf) augmented reality devices on viewing experience. We have provided an analysis of data associated with the feeling of presence and workload, showing for example that today's handheld devices are perceived as providing significantly more *Possibility to examine* an AR object. We hope the presented research is timely and many fellow researchers will benefit from a detailed analysis of how viewing devices effect AR object viewing experiences.

Acknowledgments. This research was partly funded by the BMBF (Ministry of Education and Research) in the DIGISTA project (no. 01UO1820A).

References

1. Coluccia, E., Louse, G.: Gender differences in spatial orientation: a review. *J. Environ. Psychol.* **24**(3), 329–340 (2004)
2. Geary, D.C.: Sexual selection and sex differences in spatial cognition. *Learn. Individ. Differ.* **7**(4), 289–301 (1995)
3. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv. Psychol.* **52**, 139–183 (1988)
4. Linn, M.C., Petersen, A.C.: Emergence and characterization of sex differences in spatial ability: a meta-analysis. *Child Dev.* 1479–1498 (1985)
5. Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: a presence questionnaire. *Presence* **7**(3), 225–240 (1998)
6. Wobbrock, J.O., Findlater, L., Gergle, D., Higgins, J.J.: The aligned rank transform for nonparametric factorial analyses using only anova procedures. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI 2011, pp. 143–146. ACM, New York (2011)



Ranking Robot-Assisted Surgery Skills Using Kinematic Sensors

Burçin Buket Oğul^{1,2(✉)}, Matthias Felix Gilgien²,
and Pinar Duygulu Şahin¹

¹ Department of Computer Engineering, Hacettepe University, Ankara, Turkey
bb.ural@gmail.com, pinar@cs.hacettepe.edu.tr

² Department of Physical Performance, Norwegian School of Sport Sciences,
Oslo, Norway
matthias.gilgien@nih.no

Abstract. Assessing surgical skills is an essential part of medical performance evaluation and expert training. Since it is typically conducted as a subjective task by individuals, it may lead to misinterpretations of the skill performance and hence lead to suboptimal training and organization of the surgical activities. Therefore, objective assessment of surgical skills using computational intelligence techniques via sensory data has received attention from researchers in recent years. So far, the problem has been approached by employing a classification model where a query action for surgery is assigned to a predefined category that determines the level of expertise. In this study, we consider the skill assessment problem as a pairwise ranking task where we compare two input actions to identify better surgical performance. To this end, we propose a hybrid Siamese network that takes two kinematic motion data acquired from robot-assisted surgery sensors and report the probability of the first sample having a better skill than the second one. Experiments on annotated real surgery data reveals that the proposed framework has high accuracy and seems sufficiently accurate for use in practice. This approach may overcome the limitations of having consistent annotations to define skill levels and provide a more interpretable means for objective skill assessment.

Keywords: Skill assessment · Ambient intelligence in education · Ambient intelligence in health · Robot-assisted surgery · Siamese networks · LSTM

1 Introduction

A major task in medical training is the assessment of surgical actions to grade current performance of the candidate and monitor the development of skills during training activities. These activities are usually performed manually in an operation room under supervision of expert surgeons [7]. Manual assessment, even being performed by experts, has several limitations, including subjectivity, lack of consistency and reliability.

Recent developments in computer-assisted surgery provide new opportunities to employ ambient intelligence techniques for objective skill assessment [12]. Data collected during surgery activity, either from sensory or multimedia interfaces, can serve

as platform for offline analysis of the action of surgeon post operation. Recently the machine learning community has made an effort to realize such analysis, which includes the development of computational methods that can automatically identify the surgery skill level as “expert”, “intermediate” or “novice” from surgical action data [3–5, 13–15]. The major short coming of these systems is their limited ability to predict a fixed number of predefined, possibly inconsistent, categories for expertise levels. They are unable to model skill levels between these category labels. Furthermore, they can model only overall expertise level of surgeons although it seems obvious that a surgeon’s performance may vary between different surgery action.

In two recent studies, the authors considered the problem as a task of learning to rank video recordings [2, 10] instead of assigning them into predefined labels. The study aimed to build models with wide applicability of skill determination in any domain, but algorithms were also tested for surgical skill assessment in addition to other tasks. In [2], they introduced a two-stream Temporal Segment Network to capture both the type and quality of actions. As an alternative to that [10] integrated an attention pooling and temporal aggregation mechanism to a two-stream CNN model. Skill assessment through video recordings has two main limitations. First, video data processing is time and resource inefficient, which makes it difficult to run the algorithms in conventional personal computers. Second, video can record the actions in two dimensions. This is unfortunate since tracking of trajectories and velocities can only be measured in two dimensions and important information of surgery skills is lost, if the third dimension is lacking.

In this study, we use kinematic data collected from robot-assisted surgery environment to develop a method for rank-based assessment of surgery skills. The study is considered as an emerging application of data-driven ambient intelligence in education of healthcare professionals. We introduce a novel deep learning framework based on Siamese of recurrent neural networks for pairwise ranking of motion kinematics in the form of multi-variate time-series data. We present the results of experimental evaluation of our method on real life data collected from human-controlled robot arms for surgical skill assessment. We argue that our approach provides a more interpretable and reliable view of objective skill assessment while it overcomes the limitations caused by inconsistencies in subjective skill grading scales.

2 Methods

2.1 Siamese Framework for Ranking

Given two surgical actions, m and n , with their kinematic data of \mathbf{x}^m and \mathbf{x}^n , which are in the form of multi-variate time series, the task is to determine which surgical action is performed with better quality. We denote this output by p_{mn} where;

$$p_{mn} = \begin{cases} 1 & m \text{ performs better than } n \\ 0.5 & m \text{ and } n \text{ show equal performance} \\ 0 & n \text{ performs better than } m \end{cases} \quad (1)$$

We interpret this as the probability of the first surgical action being performed better than the second. Our goal is then to train a model that minimizes the probabilistic loss in human-annotated samples for surgical skills. To this end, we introduce a novel framework based on a Siamese network of recurrent neural networks integrated with a probabilistic ranking layer, which can take the case of skill equivalence into consideration (Fig. 1).

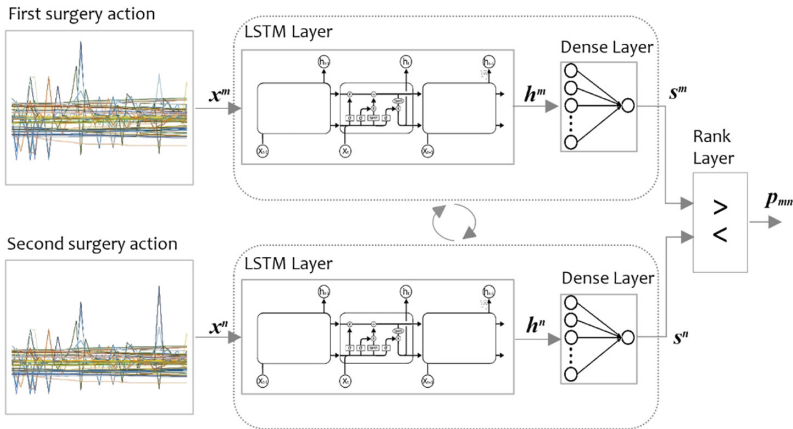


Fig. 1. Siamese network for pairwise ranking of surgery actions

LSTM Layer: The kinematic data at one input of the Siamese network is given in the form of multi-variate time-series and fed into a long short-term memory (LSTM) network [9]:

$$h_t = LSTM(h_{t-1}, x_t) \tag{2}$$

where x_t and h_t are the inputs at time t . The LSTM is parameterized by output, input and forget gates, controlling the information flow within the recursive operation. At every time step t , LSTM outputs a hidden vector h_t that reflects the skill representation of the kinematic motion at position t .

Dense Layer: A fully-connected layer takes the vector of skill representation at the output of an LSTM layer, h^m for any of the input m , and transforms it into a scalar, s^m , which is directly comparable with the output, s^n , at the other end of the Siamese network.

Rank Layer: This layer adapts a probabilistic loss function, which was originally introduced to learn how to rank text objects using gradient descent [1]. The pairwise rank between two surgery action inputs is required to be represented by p_{mn} , which is interpreted as the probability of m performing a better action than n . We denote the posterior probability distribution $P_{ij} = P(i \succ j)$, where \succ refers to the skill superiority

of i to j and let \hat{P}_{ij} be desired target values for those posteriors, such that $\hat{P}_{ij} \in \{1, 0.5, 0\}$. The goal is then to minimize the distance between these two entities.

2.2 Implementation

We used a bidirectional LSTM [8] to allow the modelling of two-way temporal dependencies in actions. The rank layer was implemented by a sigmoid activation followed by a binary cross-entropy loss function. We used the following hyper-parameters: a learning rate of 0.001, a batch size of 2 and a unit size of 64 with single hidden layer. The framework was implemented in Keras using TensorFlow back-end.

3 Results

3.1 Data

We evaluated our framework on a publicly available real surgery data set called JIGSAW [6]. The JIGSAWS dataset contains of surgical data collected from eight subjects with different skill levels performing three different surgical tasks using the da Vinci surgical system. The tasks are 4-throw suturing (39 trials), needle passing (26 trials), and knot tying (36 trials) performed on benchtop training phantoms. The dataset consists of 76 motion variables collected at 30 Hz, including tooltip positions and orientation, linear and rotational velocities, and gripper angle. Therefore, the kinematic data that we use in the study refers to a multi-variate time-series data set captured from the da Vinci robot. A trial is a part of the data set that corresponds to one subject performing one instance of a specific task. Each subject is categorized by a fixed expertise level but each trial may have a different score. This score is annotated using OSATS as a grading system [11]. OSATS consists of different grading criteria like respect for tissue, time and motion, flow of operation, overall performance and the quality of the final product. The system reports a final grade to represent general surgical skill.

3.2 Evaluation Setup

We performed four-fold cross validation to evaluate the prediction performance. In this setup, the pairs between $\frac{3}{4}$ of the actions were used for training and the remaining pairs were used for testing. Therefore, test samples include both the pairs where neither video has been used in a pair for training and the pairs where the other video was used for training in a different pairing. We used pairwise ranking accuracy, which is the percentage of correctly ordered pairs, produced by each fold. We reported two accuracy results for the cases where the skill equivalence is counted and not counted.

3.3 Empirical Results

We applied our model for each surgery task separately to rank surgery actions by their skills. Table 1 discerns the accuracy for each task where the pairs with equal skill scores are taken into account.

Table 1. Results of pairwise ranking including skill equivalence

Surgery type	Accuracy (%)
Knot tying	75.1
Needle passing	74.4
Suturing	60.3

To our knowledge, this is the first study that applies a method for pairwise ranking on kinematic data of surgery skills. Therefore, there is no previous study to benchmark our method. [2, 10] are the most relevant studies, which used video data for skill ranking and tested their methods in the same dataset. Another difference is the fact that they work for only binary ranking cases, where the equivalent skills were found to be inconsistent. To make a comparison with these methods we ran our model with complementary kinematic data from which the equally-rated pairs were removed. The results are shown in Table 2. [2, 10] did not give accuracies separately for each task, but rather reported overall performance in surgery dataset. [2] tested two different versions of their model, where only spatial information was used, and temporal data was combined with spatial data in two-streams. Our model can achieve a competitive accuracy with video-based models. Moreover, the present model is built upon only kinematic data, which reduces the computational resources compared to approaches which use videos. [2] reported that average running time to train a single fold is 18 h with NVIDIA TITANX GPU, whereas learning a fold in our model is conducted in less than an hour with a conventional CPU.

Table 2. Results of pairwise ranking excluding skill equivalence

Method	Action data	Surgery type	Accuracy (%)
Present method	Kinematic	Knot tying	79.6
		Needle passing	77.5
		Suturing	63.5
		Average	73.5
Doughty et al. 2018 (spatial)	Video	–	66.5
Doughty et al. 2018 (two-stream)	Video	–	74.4
Li et al. 2019	Video	–	73.1

4 Conclusion

We introduce a novel framework for objective skill assessment for robot-assisted surgery. The contribution of the study is twofold. First, kinematic-based surgical skill assessment problem is approached for the first time as a pairwise ranking task instead of the direct assignment of samples into predefined skill categories. This approach provides a more interpretable and reliable skill assessment while it overcomes the limitations caused by inconsistencies in subjective skill grading scales. Compared to

video-based solutions, the use of kinematic data reduces the demands on computational power and is therefore a more applicable alternative for the practical implementation in a hospital setting. Second, a novel deep learning framework based on Siamese of recurrent neural networks is introduced for pairwise ranking of multi-variate time-series data. Experimental results on surgical skill assessment data have justified the applicability of the proposed models for this task. Since the system does not rely on learning from any problem-specific features, the framework can be easily adopted for other problems in data-driven ambient intelligence with sensory interfaces.

Acknowledgments. Burçin Buket Oğul was financially supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under 2214-A program.

References

1. Burges, C.J., Shaked, T., Renshaw, E., et al.: Learning to rank using gradient descent. In: International Conference on Machine Learning, pp. 89–96 (2005)
2. Doughty, H., Damen, D., Mayol-Cuevas, W.: Who’s better? Who’s best? Pairwise deep ranking for skill determination. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)
3. Fard, M.J., Ameri, S., Darin, E.R., et al.: Automated robot-assisted surgical skill evaluation: predictive analytics approach. *Int. J. Med. Robot. Comput. Assist. Surg.* **14**(1), e1850 (2018)
4. Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.-A.: Evaluating surgical skills from kinematic data using convolutional neural networks. In: Frangi, A.F., Schnabel, J. A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 214–221. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00937-3_25
5. Funke, I., Mees, S.T., Weitz, J., Speidel, S.: Video-based surgical skill assessment using 3D convolutional neural networks. arXiv preprint [arXiv:1903.02306](https://arxiv.org/abs/1903.02306) (2019)
6. Gao, Y., Vedula, S.S., Reiley, C.E., et al.: JHU-ISI gesture and skill assessment working set (JIGSAWS): a surgical activity dataset for human motion modelling. In: MICCAI Workshop (2014)
7. Grantcharov, T.P., Bardram, L., Funch-Jensen, P., et al.: Assessment of technical surgical skills. *Eur. J. Surg.* **168**, 139–144 (2002)
8. Graves, A., Fernández, S., Schmidhuber, J.: Bidirectional LSTM networks for improved phoneme classification and recognition. In: Duch, W., Kacprzyk, J., Oja, E., Zadrozny, S. (eds.) ICANN 2005. LNCS, vol. 3697, pp. 799–804. Springer, Heidelberg (2005). https://doi.org/10.1007/11550907_126
9. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997)
10. Li, Z., Huang, Y., Cai, M., Sato, Y.: Manipulation-skill assessment from videos with spatial attention network. arXiv preprint [arXiv:1901.02579](https://arxiv.org/abs/1901.02579) (2019)
11. Martin, J., Regehr, G., Reznick, R., et al.: Objective structured assessment of technical skill (OSATS) for surgical residents. *Br. J. Surg.* **84**, 273–278 (1997)
12. Peters, B.S., Armijo, P.R., Krause, C., et al.: Review of emerging surgical robotic technology. *Surg. Endosc.* **32**(4), 1636–1655 (2018)
13. Wang, Z., Fey, A.I.: SATR-DL: improving surgical skill assessment and task recognition in robot-assisted surgery with deep neural networks. In: IEEE Conference of the Engineering in Medicine and Biology Society, pp. 1793–1796 (2018)

14. Wang, Z., Fey, A.M.: Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery. *Int. J. Comput. Assist. Radiol. Surg.* **13**, 1959–1970 (2018)
15. Zia, A., Essa, I.: Automated surgical skill assessment in RMIS training. *Int. J. Comput. Assist. Radiol. Surg.* **13**, 731–739 (2018)



uAQE: Urban Air Quality Evaluator

Claudio Rossi^{1(✉)}, Alessandro Farasin^{1,2}, Giacomo Falcone¹,
and Carlotta Castelluccio³

¹ LINKS Foundation, Turin, Italy
{claudio.rossi,alessandro.farasin,giacomo.falcone}@linksfoundation.com

² Polytechnic of Turin, Turin, Italy
alessandro.farasin@polito.it

³ Microsoft, Milan, Italy
carlotta.castelluccio@microsoft.it

Abstract. Knowing the amount of air pollutants in our cities is of great importance to help decision-makers in the definition of effective strategies aimed at maintaining a good air quality, which is a key factor for a healthy life, especially in urban environments. Using a data set from a big metropolitan city, we realize the uAQE: urban Air Quality Evaluator, which is a supervised machine learning model able to estimate air pollutants values using only weather and traffic data. We evaluate the performance of our solution by comparing the predicted pollutant values with the real measurements provided by professional air monitoring stations. We use the predicted pollutants to compute a standard Air Quality Index (AQI) and we map it into a set of five qualitative AQI classes, which can be used for decision making at the city level. uAQE is able to predict the AQI class value with an accuracy of 0.8.

Keywords: Air quality · Environment · Weather · Traffic

1 Introduction and Related Works

Air pollution is the introduction into the atmosphere of chemicals, particulates, or biological materials that causes discomfort, disease, or death to humans and to other living organisms alike. More than 5.5 million people worldwide are dying prematurely every year as a result of air pollution exposure [1]. This fact confirms that air pollution is one of the world's largest environmental health risks. Most of these deaths are occurring in rapidly developing economies, e.g., China and India, but also in European metropolitan cities, e.g., Milan or Turin, which have an air pollution index among the highest ones according to recent rankings¹.

Road transport is one of the main causes of air pollutants emissions, accounting for the 14% of the total emissions in European countries².

¹ <http://www.numbeo.com/pollution/rankings.jsp>.

² <http://www.eea.europa.eu>.

Other human activities having a strong impact on air quality are industrial processes, farming, heat and air conditioning, and other types of transport (trains, airplanes, etc.).

It is a well known fact that weather phenomena have a strong impact on air pollutants because once pollutants are emitted into the air, they propagate into the atmosphere according to weather conditions, e.g., turbulence mixes pollutants into the surrounding air, and wind carries them away from the source location. Conversely, when the air near the surface of the earth is cooler than the air above (a phenomenon called temperature inversion) there is very little air mixing. Since cool air is heavy, it will not to move up to mix with the warmer air above. Thus, any pollutants released near the surface will get trapped and build up in the cooler air layer.

Municipalities struggle to predict the effect of traffic policies, e.g., total traffic block, stop of most pollutant vehicles, on the air quality because there is a lack of easy-to-use tools that can estimate the air pollution taking into account also the meteorological predictions. Furthermore, the availability of air quality measurement stations in a city is very limited due to economic constrains. A professional station requires a non negligible investment (about 200k € per installation) and it has a high maintenance cost (about 30k € per year) [2].

Because of its importance, the estimation of the air quality has been subject to some studies. In [2], Microsoft researchers proposed a semi supervised learning approach able to predict PM_{10} and Nitrogen Dioxide (NO_2) emissions at an higher spatial resolution with respect to the one achieved by the installed air quality sensors by coupling other data sources such as traffic flows, the structure of the road network, meteorological conditions and point of interest locations. Their solution is complementary to ours, and it can be used to improve the spatial resolution of the uAQE. Other relevant studies include the [3] and [4], which present a set of learning methods able to predict NO_x concentrations from past observations and weather conditions. In [5], the authors studied Delhi's $PM_{2.5}$ concentrations and its correlation with the vehicular traffic and with the weather conditions. However, the proposed model makes several empirical assumptions and it includes parameters specific to the city of Delhi. Hence, it cannot be re-used for our purpose.

To help decision-makers in keeping under control the air quality we propose uAQE: urban Air Quality Evaluator, which is a set of supervised machine learning model able to predict air pollutants values in a urban environment using only weather and traffic data. We train our models with data taken from a big metropolitan city, i.e., Milan, building one model for each air pollutants. Our work is different from all the above mentioned approaches because we aim to predict pollutants without requiring data from air quality stations. Note that we train one model for each air pollutants, namely Nitrogen Dioxide (NO_2), Ozone (O_3), Carbon Monoxide (CO), Benzene (C_6H_6), Total Nitrogen (N_2), Particulate Matter (PM_{10}), Sulfur Dioxide (SO_2), Particulate Matter ($PM_{2.5}$), Black Carbon (BC), and Ammonia (NH_3). We present the accuracy of each model using the pollutants as measured by professional air stations. Following

a regional standard, we use the predicted pollutants to compute an Air Quality Index (AQI) which is then mapped it into a set of five qualitative classes that are used to manage air quality policies at city level. We finally asses the classification accuracy achieved by uAQE obtaining a value of 0.8.

2 Input Data

Our data has been collected in the city of Milan during two months (Nov.- Dec. 2013), and it contains three distinct data categories.

Weather: we have six different weather stations placed within the city limit. Each station has a unique ID, type, location, and it features a set of co-located sensors. Each sensor measures a different meteorological phenomena. This information has been obtained thanks to ARPA (Agenzia Regionale per la Protezione dell'Ambiente). The **weather data set** contains wind direction (degree), wind speed (m/s), temperature (Celsius degree), relative humidity (%), precipitation (mm), global radiation ($\mu W/m^2$), net radiation ($\mu W/m^2$), and atmospheric pressure (hPa).³

Traffic: through fixed video cameras already installed for traffic access control at 52 locations in the central area of Milan (*Cerchia dei Bastioni*) the local authority obtained the plate number of transiting vehicles, from which the vehicle characteristic could be extracted from the official database, i.e., the *Motorizzazione civile*, which holds the information of all Italian vehicles. Note that we received anonymized data, i.e., with hashed plate numbers and with no information about the vehicle owner. Therefore, only the technical details of each vehicle has been made available to us. These data have been provided as open data by the city of Milan. The **traffic data set** includes each vehicle passage at each gate, for which the location and the timestamp of each passage is known. For each passage, the vehicle characteristics are given, namely the European emission standard category (EURO category from 1 to 6), the vehicle type (i.e., bus, freight, transport, people transport or not available), the fuel type (i.e., petrol, diesel, electric, LPG, hybrid or missing), the presence of the Diesel Particle Filter (DPF) and the vehicle length expressed in mm.

Air: we take the measurements of three different air stations located within the city limits. Each station features multiple co-located sensors, each of which measures a single air pollutant. Also these measurements are directly provided as open data by ARPA, who is the official source of this kind of data. The **air pollution data set** contains the measured values of the aforementioned pollutants: NO_2 ($\mu g/m^3$), NH_3 ($\mu g/m^3$), C_6H_6 ($\mu g/m^3$), SO_2 ($\mu g/m^3$), BC ($\mu g/m^3$), CO ($\mu g/m^3$), N_2 (ppb), PM_{10} ($\mu g/m^3$), $PM_{2.5}$ ($\mu g/m^3$), O_3 ($\mu g/m^3$).

We compute the hourly air quality index as defined by Piedmont index AQI because is the only example of operational use of an air quality index in Italy⁴. The AQI uses only three pollutants, namely NO_2 , PM_{10} , O_3 , and it is formulated as $I_{AQI} = \frac{I_{PM_{10}} + \max(I_{NO_2}, I_{O_3})}{2}$, where $I_{PM_{10}} = \frac{V_{avg24h_{PM_{10}}}}{V_{ref_{PM_{10}}}} \times 100$,

³ http://ita.arpalombardia.it/ITA/qaria/doc_RichiestaDati.asp.

⁴ <http://www.arpae.it/cms3/documenti/aria/IQA.pdf>.

$I_{NO_2} = \frac{Vmaxh_{NO_2}}{Vref_{NO_2}} \times 100$, $I_{O_3} = \frac{Vmax8h_{O_3}}{Vref8h_{O_3}} \times 100$. $Vref$ are reference values, while $Vavg24h$, $Vmaxh$, $Vmax8h$ means values averaged over the last 24 h, hourly maximum, maximum over the last 8 h, respectively.

We observe that in the considered data set O_3 never exceeds the maximum value established for preserving human health (i.e., $Vref8h_{O_3} = 120 \mu\text{g}/\text{m}^3$), whereas NO_2 exceeds its hourly maximum value (i.e., $Vref_{NO_2} = 200 \mu\text{g}/\text{m}^3$) only in few cases (<5%). Conversely, PM_{10} exceeds the daily maximum value (i.e., $Vref_{PM_{10}} = 50 \mu\text{g}/\text{m}^3$) in 50% in the cases.

We map the computed AQI in the five classes defined by the Piedmont region, namely Optimal ($0 \leq AQI < 50$), Good ($50 \leq AQI < 75$), Fair ($75 \leq AQI < 100$), Average ($100 \leq AQI < 125$), Not Very Healthy ($125 \leq AQI < 150$), Unhealthy ($150 \leq AQI < 175$), Very Unhealthy ($AQI \geq 175$). In our data set, we observe that there are no AQI values in the Optimal level and very few in the Good one, while the most part of values ($\approx 80\%$) falls between Fair and Not Very Healthy levels.

3 Feature Construction

Our aim is to use a supervised machine learning approach to predict pollutants from traffic and weather data under the hypothesis that we do not have air sensors to directly measure air pollutants.

We merge the three data categories previously described (weather, traffic, air) at hourly resolution because this is the maximal temporal resolution of both weather and air data. We average the measurements produced by different sensor of the same type in the same hour, while we count the hourly passages at all gates of vehicles. Specifically, we count the total hourly passages of each EURO category, type, fuel, DPF availability. In order to perform all data manipulations, we use *R* and the *plyr* library, which provides data aggregation operators. We fill few missing values (<1%) in the air and weather data by polynomial interpolation using the spline function of the *zoo* library.

The final feature set is composed by the following variables:

- **Time:** day of week (1–7), hour (1–24). This is to consider the regular patterns of human activities, which are framed within the day and within the week;
- **Hourly passages:** counts of total passages and aggregated count by EURO class, vehicle type, fuel type, existence of particulate filter. Because we compute the total passages, we remove one category from each aggregation to avoid creating features which are linear combination of other ones while reducing the number of total features;
- **Hourly weather phenomena averages:** wind direction, wind speed, temperature, relative humidity, precipitation, global radiation, net radiation, atmospheric pressure.

Additionally, in order to consider the effect of the past on the current pollutants levels, for each traffic and weather feature $f(t)$ (wind direction excluded) we add another feature $fp(t)$ that at each time instant c is equal to the sum of

$f(t)$ over the last x hours ($fp(t) = \sum_{t=c-x}^{t=c} f(t)$). We evaluate increasing values of x starting from 1 and we empirically find the best value to be 12. Studying the correlation between weather and pollutants we notice that temperature, relative humidity, precipitation wind speed, and atmospheric pressure are the most significant ones, having an average absolute correlation of 0.40, 0.20, 0.13, 0.51, 0.51 with the pollutants considered in the AQI computation, respectively. Conversely, all traffic features results less correlated and they are not shown for brevity.

Table 1. Performance comparison between the GLM and the BRNN reporting average absolute error and its standard deviation for each pollutant.

Agent	Unit	BRNN 9 Neur.		GLM		Comparison	
		$\mu(\varepsilon)$	$\delta(\varepsilon)$	$\mu(\varepsilon)$	$\delta(\varepsilon)$	$\Delta[\mu(\varepsilon)]$	$\Delta[\delta(\varepsilon)]$
NO_2	$\mu g/m^3$	12.45	10.28	21.02	20.58	41%	50%
O_3	$\mu g/m^3$	22.47	20.06	46.47	45.09	52%	56%
CO	$\mu g/m^3$	10.99	10.18	19.31	15.59	43%	35%
C_6H_6	$\mu g/m^3$	39.85	64.44	98.92	145.89	60%	56%
N_2	ppb	27.33	29.12	56.27	81.32	51%	64%
PM_{10}	$\mu g/m^3$	13.65	14.35	34.56	34.62	61%	59%
SO_2	$\mu g/m^3$	18.64	20.67	38.58	42.68	52%	52%
$PM_{2.5}$	$\mu g/m^3$	13.28	13.88	32.55	31.87	59%	56%
BC	$\mu g/m^3$	18.47	24.21	35.91	66.30	49%	63%
NH_3	$\mu g/m^3$	26.95	52.95	41.83	119.12	36%	56%

4 Pollutants Prediction and and Evaluation of AQI

We implement the machine learning models using the *caret* package. In particular, we test several machine learning algorithms for regression, including the Generalized Linear Model (GLM), the Random Forest (RF), the Support Vector Machines (SVM), and the Artificial Neural Networks (ANN). We test all algorithms with default hyper-parameters and with the same random seed, uniformly selecting in time the 70% of the samples as training set, and leaving the remaining 30% for the test set. We train all models with a 5-fold cross-validation and we compute the model performances in terms of mean squared error for pollutants (regression problem). For brevity, we report only the results of the GLM, which we consider as the baseline, and of the ANN, which is the model that performs best.

Artificial Neural Networks (ANNs) [7] are inspired by biological nervous systems such as the human brain, which process information through a large number of highly interconnected processing elements (neurones). ANNs can be used in

several applications, such as pattern recognition or data classification, and they are a supervised machine learning technique.

Specifically, we choose a particular type of ANN, namely the BRNN model (Bayesian Regularization of Neural Networks) [6], because it is more robust than standard back-propagation networks and it can reduce the need for lengthy cross-validation. The main model parameter is the number of neuron n to be used. In order to define the optimal n , we incrementally evaluate the model accuracy starting with $n = 1$ and incrementing it in steps of 1 until 20. Therefore, we empirically find the best value of $n = 9$, after which the performance improvement can be considered negligible.

For each pollutant, we compare the BRNN model with the GLM performances, obtaining for the BRNN an improvement of the average absolute error between 36% and 61% over the GLM. The performance comparison is fully reported in Table 1.

Using the predicted values of NO_2 , O_3 and PM_{10} , we compute the AQI value, and then map into the classes described in Sect. 2 (i.e. Optimal, Good, Fair, Average, Not Very Healthy, Unhealthy, Very Unhealthy).

Our model predicts AQI with a class accuracy of 0.8, which we evaluate as satisfactory, especially considering that the distance of the classification error is never greater than one, meaning that when the model predicts an erroneous class it is never beyond the adjacent one, e.g., the model can predict Fair instead of Good but it never predicts Average or any worst condition instead of Good.

5 Conclusion and Future Works

In this paper we used traffic and weather data in order to predict the air pollution in a metropolitan city. We designed and implemented a set of machine learning models to predict single pollutants that we used to compute a qualitative air quality classed based on a standardized Air Quality Index (AQI). The performance of our best model, (BRNN with 9 neurons), achieves an AQI class accuracy of 0.8.

Future works will include the evaluation of our approach on a bigger dataset, an improvement of the feature set, and the evaluation of several scenarios (e.g., including partial or complete traffic block, different weather conditions, etc.) in order to evaluate the impact of local traffic policies on the air quality.

References

1. Amos, J.: Polluted air cause 5.5 million deaths a year new research says. BBC NEWS, Science and Environment (2016)
2. Zheng, Y., Liu, F., Hsieh, H.: U-air: When Urban Air Quality Inference Meets Big Data. In: Microsoft Research Asia. ACM (2013)
3. Juhosa, I., Makrab, L., Ttha, B.: Forecasting of traffic origin NO and NO2 concentrations by support vector machines and neural networks using principal component analysis. *Simul. Model. Pract. Theory* **16**(9), 1488–1502 (2008)

4. Berkowicz, R., Palmgren, F., Hertel, O., Vignati, E.: A study on effects of weather, vehicular traffic and other sources of particulate air pollution on the city of Delhi for the year 2015. *J. Environ. Pollut. Hum. Health* **4**(2), 24–41 (2016)
5. Gopaldaswami, R.: Using measurements of air pollution in streets for evaluation of urban air quality meteorological analysis and model calculations. *Sci. Total Environ.* **189–190**, 259–265 (1996)
6. Burden, F., Winkler, D.: Bayesian Regularization of Neural Networks. PubMed (2008)
7. Stergiou, C., Siganos, D.: Neural networks. Imperial College London (2011)



Enhancing an Eco-Driving Gamification Platform Through Wearable and Vehicle Sensor Data Integration

Christos Tselios¹(✉), Stavros Nousias¹, Dimitris Bintzas¹,
Dimitrios Amaxilatis², Orestis Akrivopoulos², Aris S. Lalos^{1,4},
Konstantinos Moustakas¹, and Ioannis Chatzigiannakis³

¹ University of Patras, 26500 Achaia, Greece

{tselios,nousias,bintzas,aris.lalos,moustakas}@ece.upatras.gr

² SparkWorks ITC Ltd., Alticham, UK

{d.amaxilatis,akribopo}@sparkworks.net

³ Sapienza University of Rome, Rome, Italy
ichatz@diag.uniroma1.it

⁴ Industrial Systems Institute, ATHENA Research Center, Patras, Greece

Abstract. As road transportation has been identified as a major contributor of environmental pollution, motivating individuals to adopt a more eco-friendly driving style could have a substantial ecological as well as financial benefit. With gamification being an effective tool towards guiding targeted behavioural changes, the development of realistic frameworks delivering a high end user experience, becomes a topic of active research. This paper presents a series of enhancements introduced to an eco-driving gamification platform by the integration of additional wearable and vehicle-oriented sensing data sources, leading to a much more realistic evaluation of the context of a driving session.

Keywords: Gamification · Eco-driving · Sensors

1 Introduction

Road transportation greatly aggravates environmental pollution given the fact that approximately 30% of the total CO₂ emissions worldwide derive from internal combustion engines used in motorized vehicles [1]. Abolishing highly polluting engines and replace every car with an electric one is probably the proper method for tackling pollution, however this will not happen overnight. It is therefore essential to motivate drivers towards adopting a more eco-friendly driving style for actually having a chance of achieving any substantial ecological or financial benefit. Alas, an individual's driving style is fused by a series of factors often

Part of this work has been supported by H2020-ICT-24-2016 Project GamECAR (Grant No. 732068) and H2020-SC1-DTH-2018-1 Project SmartWork (Grant No. 826343).

including psychological, cultural and societal ones which are difficult to identify as well as change. Yet, gamification proved to be an effective tool towards guiding targeted behavioural changes and focusing on such methods could be something worth considering.

Gamification is directly correlated with the introduction of specialized, game-related guidelines utilized to facilitate the subconscious adoption of a desired behaviour. Users are encouraged to follow specific ways of dealing with issues through the stimulation of intrinsic personal traits, often referred as *core drives* [2]. The stimulation of core drives such as (i) accomplishment, (ii) empowerment of creativity, (iii) ownership, (iv) relatedness, (v) scarcity, (vi) unpredictability and (vi) curiosity, greatly increase user interest, commitment and involvement towards concluding any given task. Individuals are motivated and essentially guided to adopt pre-defined execution patterns via a playful experience, specifically designed to elevate awareness and negative outcome provision ability, all through participation in a seemingly simplistic game.

As stated in [2] the main pillars of gamification combine a large variety of verticals such as technological platforms, user experience and game mechanics optimization, motivational psychology and behavioural economics, all glued together through an underlying business logic. Following a more person-oriented approach, Kim et al. [3] suggested differentiating between intrinsic and extrinsic rewards while Chou proposed the *Octalysis* actionable gamification framework [2] which incorporates previous approaches in an attempt to realistically determine elements which make games both meaningful and fun.

It becomes obvious that eco-driving could be significantly improved through a gamification platform which will guide users towards adopting a more environmental friendly driving attitude. The rest of the paper is organized as follows: Sect. 2 presents some of the elements for designing realistic eco-driving gamification platform while Sect. 3 focuses on how accumulating data from vehicle and wearable sensors boosts the platform's realism and further engages users. Section 4 describes the prototype implementation process and finally, Sect. 5 draws conclusions and summarizes the paper.

2 Gamification Framework Prototype for Improving Eco-Driving Behaviour

The original prototype of the gamification platform was developed primarily to motivate users to adopt an eco-friendly driving style through real-time push notifications and auto-generated feedback. Safe driving was not to be compromised in any way therefore all platform-oriented messages were designed to be non-intrusive allowing drivers to focus on the road. The *Octalysis* actionable gamification framework was taken into consideration, yet without neglecting complementary gamification methodologies, each with significant benefits. All playful interventions were adapted to the characteristics of the driver by a dedicated module that was based on cutting-edge user models for expressing one's driving behaviour.

The interaction between users and the gamification platform was designed to be made via their real driving sessions, the tracking of which was rendered possible by dedicated equipment carried on board. The cornerstone of this equipment was the driver’s smartphone and the variety of features it integrates, such as the accelerometer and the GPS Navigation system. Additional, more direct parameters related to one’s specific driving skills such as hard/soft braking as well as an overall aggressiveness score linearly linked to the total number of braking/accelerating actions per minute, were calculated, associated with the driver’s profile and stored for future reference. Figure 1 presents a flowchart of the overall process which allowed drivers to improve their eco-driving skills through serious gaming [4]. All elements are tailor-made to automatically align with the various levels of player experience which progressively increases as the game progresses, while each level has distinctive characteristics, features and objectives.

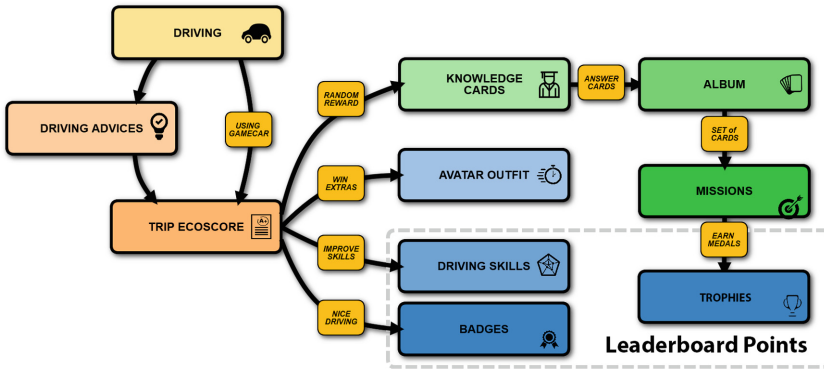


Fig. 1. Gamification flowchart designed to improve eco-driving behaviour

As shown in Fig. 1, each driving session mainly contributes to the creation of the *Trip Ecoscore* which is the most important gamification mechanism. It provides information to the user regarding the eco-driving quality of the previous session and is immediately correlated to the overall progress. Moreover, it acts as the basis for every possible reward generation, thus is directly linked with the core drives of the *Octalysis* gamification framework [2]. Obtaining an acceptable Trip Ecoscore is the key for (i) earning *Badges*, (ii) collecting points which improve the overall Driving Skills, (iii) be rewarded with access to *Knowledge Cards* which unlock *Missions* and consequently *Trophies*. Badges, Driving Skill score and Trophies are visible to the platform’s *Leaderboard*, an online *Hall of Fame* where eco-driving record scores are kept. In addition, each user owns an *Avatar* acting as their virtual representation, starting from a basic avatar and gradually unlocking multiple avatar outfit parts during their progression in the game. The avatar is displayed in the driver’s *Profile*, a crucial gamification element as it provides a clear view of a player’s progression in all aspects of the game.

3 Increasing Realism Through Sensor Data Assembly

Despite the overall functionality of the implemented prototype, some issues were soon identified. Relying on the smartphone's accelerometer for getting a rough estimation of the vehicle's driving condition proved to be an inaccurate approach. The system failed to expose the significant variations in cruising speed when those occurred after smooth acceleration or braking, thus returned an erroneous Trip Ecoscore value estimation. Fast moving vehicles were treated the same way with slow moving ones, when speed remained consistent for a longer duration of time, a behaviour which does not comply with the real consumption and CO₂ emission. This situation was tackled with the introduction of additional metrics obtained from two additional categories of sensing devices each contributing with specific bits of information for filling the gaps and reproduce the context of each route in a realistic manner. A high level architecture of the new Mobility Monitoring Network prototype is presented in Fig. 2.

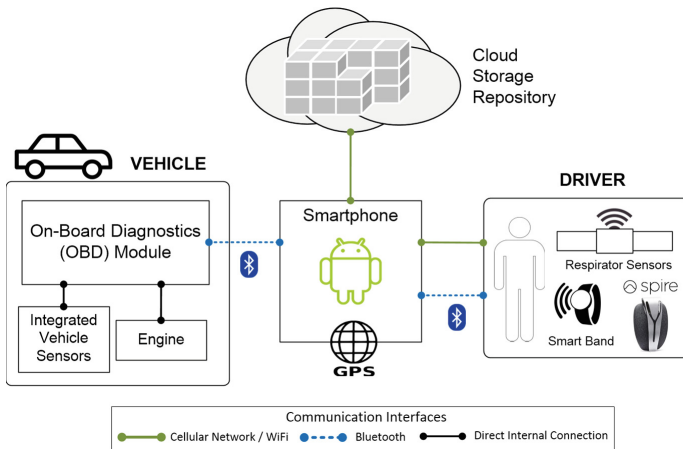


Fig. 2. High level architectural diagram of the Mobility Monitoring Network prototype

For implementing a driving behaviour monitoring framework which will reveal the condition of both the vehicle as well as the driver, it is essential to record a series of physiological, behavioural, environmental and vehicle parameters. Therefore, input was obtained from two major sensing categories, vehicle and driver-oriented.

1. **Vehicle sensors** measuring gear shifts, tire pressure, temperature and oil are already integrated to modern cars by virtually every manufacturer, linked through an internal controller area network (CAN) for ensuring robust communication. Data can be retrieved using an on-board diagnostics (OBD) controller and reveal a detailed log of the vehicle's condition on any given time.

2. **Wearables** are attached to the driver's body and non-intrusively record aspects of his physiological condition. The obtained traces (raw data) are encrypted and stored in the internal memory of the device remaining available for further process.

The smartphone retained the role of the main coordination node due to the large number of communication interfaces, caching and processing capability it supports as well as the auxiliary data sources it incorporates. The overall functionality of establishing communication with all necessary devices, retrieve data and process them accordingly was integrated into the gamification application of the original prototype. Moreover, techniques for data-preprocessing, customized for extracting the most significant information [5] were also integrated. For the preliminary prototype evaluation and without the loss of generality, only data regarding vehicle speed, engine rounds-per-minute (RPM), braking, throttle position and driver heartbeats per minute were collected for creating matrices stored in a per-trip fashion. The smartphone acted as a data aggregator that accumulated sensor values, added a timestamp and created a .CSV file. This file was also populated by additional content retrieved from the smartphone's GPS which indicated the exact positioning of the driver/vehicle. At the end of each route, the .CSV file was transmitted to the Cloud Storage Repository using affiliated WiFi or 4G/LTE network connection.

4 Gamification Accuracy

Unlike the original prototype which only relied on acceleration and speed changes, it is now possible to obtain additional and highly accurate information for braking, RPM and throttle position. This lead to a more precise eco-score function which penalizes high level of engine RPM, is associated with gear shift-up while cruising and subtracts points for cases of abrupt braking and high acceleration, in correlation with the driver's heartbeat condition. Moreover, a complementary metric, the aggressiveness score penalizes high lateral acceleration, abrupt braking and high variances in throttle position and RPM. Both metrics contribute to the Trip Ecoscore value which remains crucial for every stage of the game. Eco-score is comprised of five parameters, shift-up parameter, braking parameter, acceleration parameter, RPM parameter and cruising parameter. All parameters are computed within a 30s temporal window. For each metric, we utilize a sigmoid function $\sigma(x)$ and a weighted histogram function that assigns each event to a bin depending on the intensity of the event. $\sigma(x)$ is defined as

$$\sigma(x) = a_2 + \left(\frac{a_1}{a_4 + e^{(a_3 \cdot [x - x_0])}} \right) \quad (1)$$

where parameters a_1 to a_4 are experimentally defined.

The aggressiveness score calculation is based on parameters such as (i) engine RPM variance, (ii) braking intensity and (iii) acceleration magnitude.

As expected, high variance of engine RPM, especially when paired with higher than normal heartbeat rate, indicates driver nervousness and improper vehicle handling. In most cases, significant fluctuations in engine RPM occur when drivers are not focused on smooth cruising and overall safe driving but tend to constantly accelerate, thus minimizing their fuel efficiency, eco-friendliness and consequently the overall road safety. Given a temporal window of $t = 30$ s, the aggressiveness score AG_{RPM} derived only from engine RPM fluctuation is

$$AG_{RPM} = \frac{s_{RPM}^2}{\mu} \quad (2)$$

where s_{RPM}^2 is the engine RPM variance and μ is an experimentally defined parameter. Another metric not available in the first prototype is Braking Intensity BI , calculated by correlating vehicle deceleration and braking duration, during the online analysis phase. This metric cumulatively contributes to the computation of aggressiveness score given the fact that abrupt braking events indicate both poor planning and low level of situation awareness regarding road conditions from the driver's side. The braking intensity aggressiveness score BI_{AG} is equal to

$$BI_{AG} = \frac{B_a}{B_a + B_s} \quad (3)$$

where B_a is the number of abrupt braking events and B_s is the number of smooth braking events within the aforementioned temporal window.

5 Conclusions

This work presented an ongoing effort to improve a fully functional eco-driving gamification platform prototype by integrating additional datasets obtained by vehicle and wearable sensors. The introduction of a holistic and highly accurate data retrieval platform greatly optimizes the overall in-game experience, which is now directly linked and fully correlated with the actual driver, vehicle and route conditions. Preliminary results indicate that drivers were able remain actively engaged in all gaming activity, while the ultimate goal of endorsing a more eco-driving behaviour is achieved.

References

1. Santos, G.: Road transport and CO2 emissions: what are the challenges? *Transp. Policy* **59**(1), 71–74 (2017)
2. Chou, Y.-K.: *Actionable Gamification: Beyond Points, Badges, and Leaderboards*, 2nd edn. Octalysis Group, Fremont (2015)
3. Kim, A.-J.: Gamification 101: design the player journey. *Gamification WS* **1**(1) (2011)
4. Nousias, S., et al.: Exploiting gamification to improve eco-driving behaviour: the GamECAR approach. *Electron. Notes Theor. Comput. Sci.* **343**, 103–116 (2019)
5. Nousias, S., et al.: Managing nonuniformities and uncertainties in vehicle-oriented sensor data over next generation networks. In: *IEEE PerCom Workshops*, Athens, pp. 272–277 (2018)



A Distributed Multi-Agent System (MAS) Application For continuous and Integrated Big Data Processing

Ariona Shashaj^(✉), Federico Mastroilli, Massimiliano Morrelli,
Giacomo Pansini, Enrico Iannucci, and Massimiliano Polito

Network Contacts, Molfetta, BA, Italy

{ariona.shashaj,federico.mastroilli,massimiliano.morrelli,
giacomo.pansini,enrico.iannucci,massimiliano.polito}@network-contacts.it

Abstract. With the advent of Ambient Intelligence (AmI) as an interdisciplinary methodology which ranges from Ubiquitous and Pervasive computing to Artificial Intelligence with the final aim to create a sensitive and responsive environment, the focus is moving towards integrated solutions of the encompassed technologies. Multi-Agent System (MAS) approach, characterized by a set of autonomous intelligent agents which can cooperate in order to achieve a common goal, can help the development of AmI integrated solutions. In this paper, we present a distributed MAS environment, *Multi-Agent Specialized system* (MASs), which supports the development of integrated AmI solutions. An application scenario considering the case of continuous Big Data processing is shown.

Keywords: Multi-Agent system · Distributed system · Big Data Processing.

1 Introduction

Ambient Intelligence (AmI), has emerged as an inter-disciplinary domain which encompass technologies in the area of Ubiquitous and Pervasive computing, Human Computer Interactions, Internet of Things and Artificial Intelligence, and is advancing towards the development of an intelligent environment sensitive and responsive to human interaction [2]. Application scenarios of AmI have been exploited in the area of smart house/cities. In the *Telco* domain, AmI applications can support the automation processes of *Customer/Citizen Operations* (CO), where human operators/users need to interact with virtual ‘bots’ (chatbot, conversational AI) in order to tackle CO tasks. In this scenario, the efficiency of dialog management is crucial and it can be enhanced by accessing a Knowledge Base which harvest from continuous Big Data streams processing.

Several works exploit the adaption of distributed Multi-Agent System (MAS) application in the domain of AmI [14, 16]. The heterogeneous, distributed and self-adaptive nature of MAS makes this approach a key enabler for the development of complex integrated solution in dynamic environments. In this paper,

we present *Multi-Agent Specialized system* (MASs), a distributed MAS environment capable to support the complete development and deployment cycle of multi-agent applications. Furthermore, we discuss an application scenario in the domain of continuous Big Data processing. The remainder of the paper is organized as follows. In Sect. 2 we motivate our work and review related works. MASs environment is presented in Sect. 3, while Sect. 4 shows a case study of continuous Big Data processing over MASs. Finally, conclusion remarks and future works are given in Sect. 5.

2 Motivation and Related Works

Our study was motivated by the demand of developing a heterogeneous Big Data oriented environment able to support *Customer/Citizen Operations* (CO), which are typically involved in the *Telco* domain. The final goal is to define an open, hybrid and interconnected architecture which allows seamless cooperation between agents and human operator. On the other hand, a Big Data oriented model is necessary in order to efficiently process long term stored data, as well as real-time ones. The Big Data Analytic domain is characterized by a set of properties, named the “5V”, which pose many challenges: **Volume** (large datasets with high complexity); **Velocity** (a fast growing stream of continuously new generated data, which need to be processed real-time or at least while they’re still significant); **Variability** (different type of data within a different structure); **Value** (usefulness of the information embedded in the raw data); **Veracity** (reliability attributed to the data) [9]. Data variability in both structure and value over time is becoming more demanding, resulting in new data mining techniques such as *stream mining*, which typically process the continuous flow of data by sampling it. This can results in scenarios where valuable data are not considered [13]. New general techniques in order to manage this growing variability of Big Data applications (a very large and continuous flow of data different in structure and value) are necessary [3, 19]. MAS paradigm defines a system composed by autonomous software which can cooperate in order to achieve a common goal. Further, MAS applications are characterized by a set of properties, named *self-properties* (*self-organization*, *self-configuration* etc), which are key attributes on defining complex and dynamic systems [7]. In the past decades, an increasing number of MAS platforms and environments have been developed, discriminated by the design principles: agent-centric (focused on reasoning architecture) [5, 15] and middleware-centric (focused on MAS environments as a collection of software which manage system processes) [4, 18]. What emerged is the adaptation of a *de-facto* standard for agent management and interaction processes, in order to guarantee interoperability between different MAS (FIPA [1]). What is missing is an environment able to manage the whole development and deployment cycle of hybrid distributed multi-agent applications. The coexistence between MAS approaches and Big Data Analytic has been considered in earlier works [6]. In [17], a MAS model has been proposed for Big Data Lambda Architecture (LA) [10], where layers of the LA are served through agents with different role.

The authors propose: *Archiver* agent (which handles incoming data and passes it to the data storage, e.g. the *Hadoop Distributed File System HDFS*), *Batch Driver* agent and a pool of *Batch Worker* agents to handle processes involved in the Batch Layer; *Stream Processing* agent and a pool of *Real-Time Worker* agents to handle Speed Layer processes; *Batch Aggregator* agent, *Real-Time Aggregator* agent and *Service* agent which handle operations of the Serving Layer. Benefits of a Multi-Agent Lambda Architecture (MALA) for Big Data Analytic, such as incremental learning, smaller cluster configuration and improvements on training time, considering the e-commerce scenario, have been exploited in [12]. In [3] Adaptive Multi-Agent System (AMAS) architecture have been applied to Big Data Analytic pipeline processes. The main idea is to split data stream process between multiple agents which are in a self-adaptive multi-feature relationship with each other.

3 Multi-Agent Specialized System (MASs)

Multi-Agent Specialized system (MASs), is a FIPA-compliant MAS environment which supports the development and deployment cycle of hybrid distributed multi-agent applications, where autonomous virtual agents and humans cooperate in order to achieve a common task. The actors of applications developed in MASs are Agents (autonomous entities able to sense and interact with the environment) and Behaviours (entities encapsulating the behavioural logical flows associated with agents). Multiple Behaviours can be associated to a single Agent, whereas their execution is triggered by the application demands. On the other hand, a single Behaviour type can be associated to different agents.

Agents are extensions of the following super-types: *Dynamic Agent* is an entity instantiated based on application demand, which execution stops on request fulfillment; *Static Agent* is instantiated based on application/system demand, its execution is stopped only when requested; *Session Agent* is an entity which is executed during a time frame (session time); *Instanceable Agent* instantiation can be triggered by other agent; *System Agent* is an entity which provides system management operations. Behaviours are classified as: *Synchronous Behaviour*, where its execution flow is tackled in a synchronous manner; *Asynchronous Behaviour* flow are executed through a dedicated thread, in an asynchronous fashion. The distributed MASs platform (execution environment of MASs agents) consists of a set of Containers, which are executed on single machine nodes. The *Main Container* is the coordinator node. In order to assure robustness against node failures, *Main Container* is re-elected through a distributed leader election algorithm, revisited from [8], and which has a linear computational cost. As it is recommended by the FIPA standard, there are two special *System Agents* executed in MASs platform: AMS (Agent Management System), which represents a primary management entity, and DF (Director Facilitator), which maintains a list of available service provided by other Agents.

FIPA interaction protocols are implemented in order to manage agents interactions. Currently, MASs supports FIPA-Request (a one-shot protocol where a

request is followed by a response message containing results or failure notification), FIPA-Subscribe (a multiple interaction protocol where a request is followed by a continuous stream of responses) and FIPA-Contract-Net (a multiple participant protocol designed to support cooperation and negotiation processes among agents). Interaction protocols are built on top of message transport standards commonly used in IoT domain (REST, MQTT).

Despite the distributed platform, MASs environment contains a set of satellite applications software, in order to support the process of agent development and deployment. In particular, MASs DAD (Distributed Agent Developer) is a graphical web application dedicated to design processes related to Agent and Behaviour implementation, whereas MASs Console is the web application which provide a continuous monitoring of the entire environment.

4 Application: Big Data Processing over MASs

Having MASs interact with the Big Data Framework (a collection of components involved in Big Data processing) proves very efficient for a number of reasons: agents provide versatility in analyzing (real time) data of heterogeneous sources and compositions; simplify modeling complex architectures for data processing (basically distributing every specific task of a pipeline to a different agent); guarantee simple integration of new software technologies (instead of rewriting entire portions of a classic software, simply integrate in MASs a new agent deputy to manage the new features). Accessing the Big Data Framework guarantees the chance to a deep and ever-continuing training over a ever growing knowledge base. We introduced pre-processing and training over a knowledge base to build a pipeline of agents for Natural Language Processing. The distributed nature of MASs, gave us the chance to exploit clusterization in order to speed up exponentially pre-processing and training jobs (see Table 1).

Table 1. Comparison of the dataset building time between single machine environment and Spark distributed environment. Single machine environment: Intel i5-3570 @ 3.40 GHz QuadCore, 8 GB RAM, 250 GB SSD, Linux Ubuntu 16.04 LTS 64 bit. Distributed environment: cluster of 9 homogenous VMWARE virtual machines(MB Intel 440BX, Intel Xeon Gold 6140@2.30 GHz DualCore, 8 GB RAM, 140 GB 53c1030 PCIe-X Fusion-MPT HDD).

Dataset to analyze	Single machine environment	Distributed environment
Single terms dictionary	3 h 20 min	2 min 44 s
2-grams dictionary	10ys (empirically)	6 min 13 s
3-grams dictionary	Out of Memory	9 min 43 s
1+2+3 combination	Out of Memory	12 min 27 s

At the same time we had to reverse engineer sequential machine learning algorithms in a new distributed fashion, in order to efficiently execute them in

©Apache Spark framework, which contains different components necessary for processing Big Data. In particular Spark Streaming opens (and keeps up) a communication channel which contains the Spark Job; Apache Kafka is responsible for the data stream interaction between MASs and the Big Data Framework; MLlib gives us a library to implement Machine Learning methods (in addition to which there's a K-nearest Neighbours model implemented in-house by our Software Engineering team [11]), while we exploited DeepLearning4J tools to elaborate Deep Neural Networks. In order to interact and efficiently exploit the functionalities of the external software which compose the Big Data Framework we developed a set of agents: *Launcher Agent* to launch a Spark job/Spark Streaming for the creation of the collections which constitute the Knowledge Base or in order to load the models, train the semantic (NLP related) agents and interacts with the user request operating on the Kafka environment; the *Producer* and *Consumer* agents respectively load the necessities infos to elaborate the job on a *Topic* and read the (eventual) result from another topic. There's not one specific pipeline, but every agent is arranged to interact with the others of the set (Fig. 1).

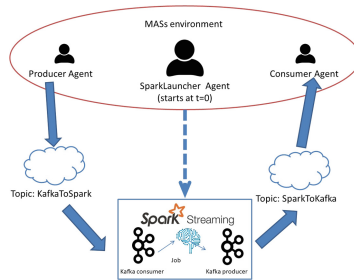


Fig. 1. Simple schema of how agents may approach the BigData framework. The *Producer* and *Consumer* agents can interact with other non-BD related agents. *Spark-Launcher* is a standalone startup agents (initiate when the MASs is boot) only responsible to start the Spark Stream/Job. The Producer and Consumer agents may even overlap in one single agent having a Producer and Consumer *behaviour*.

5 Conclusion and Future Work

The growing interest on Ambient Intelligence (AmI) has pointed out the need to develop technical integrated solutions, as well as of frameworks and environments which support their design and execution process. In this paper, we present MASs (Multi-Agent Specialized system), a distributed multi-agent system environment able to support the development end execution of integrated AmI applications. We have discussed a practical application where continuous Big Data processing has been managed through MASs.

This is an ongoing work. Future scenarios will focus on the application of the complete MASs+BigData environment in order to optimize Customer/Citizen

Operations: we intend to realize working Intelligent Virtual Agents environment able to interact with customer in Natural Language (requiring deep Natural Language Processing and self-training capabilities) and to process data coming from such interactions to extract hidden trends, as well as, perform sentiment analysis. Moreover, the input data types will include other data source types, such as vocal speech interactions.

Acknowledgments. Funding/Support: This work was supported by the Horizon 2020-PON 2014/2020 project B.4.M.A.S.S “Big Data for Multi-Agent Specialized System”.

Contribution: The MASs environment has been developed by *Ingegneria dei Sistemi* Department, Network Contacts, Molfetta, Italy.





References

1. The foundation for intelligent agent. http://www.fipa.org/repository/standard_specs.html
2. Augusto, J.C., McCullagh, P.J.: Ambient intelligence: concepts and applications. *Comput. Sci. Inf. Syst.* **4**(1), 1–27 (2007)
3. Belghache, E., Georgé, J.P., Gleizes, M.P.: Towards an adaptive multi-agent system for dynamic big data analytics. In: 2016 International IEEE Conferences on Ubiquitous Intelligence & Computing (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), pp. 753–758. IEEE (2016)
4. Bellifemine, F., Caire, G., Poggi, A., Rimassa, G.: Jade: a software framework for developing multi-agent applications lessons learned. *Inf. Softw. Technol.* **50**(1–2), 10–21 (2008)
5. Bordini, R.H., Hübner, J.F., Wooldridge, M.: *Programming Multi-agent Systems in AgentSpeak using Jason*, vol. 8. Wiley, Hoboken (2007)
6. Cao, L., Gorodetsky, V., Mitkas, P.A.: Agent mining: the synergy of agents and data mining. *IEEE Intell. Syst.* **24**(3), 64–72 (2009)
7. Di Marzo Serugendo, G., Gleizes, M.P., Karageorgos, A.: *Self-organising software: from natural to artificial adaptation* (2011)
8. Garcia-Molina, H.: Elections in a distributed computing system. *IEEE Trans. Comput.* **1**, 48–59 (1982)
9. Jagadish, H., et al.: Big data and its technical challenges. *Commun. ACM* **57**(7), 86–94 (2014)
10. Marz, N., Warren, J.: *Big Data: Principles and Best Practices of Scalable Real-time Data Systems*. Manning Publications Co., New York (2015)
11. Morrelli Massimiliano, C.M.: Nuova frontiera della classificazione testuale: big data e calcolo distribuito, pp. 1–20 (2019)
12. Pal, G., Li, G., Atkinson, K.: Multi-agent big-data lambda architecture model for e-commerce analytics. *Data* **3**(4), 58 (2018)
13. Palpanas, T.: Big sequence management: a glimpse of the past, the present, and the future. In: Freivalds, R.M., Engels, G., Catania, B. (eds.) *SOFSEM 2016*. LNCS, vol. 9587, pp. 63–80. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49192-8_6
14. Piette, F., Caval, C., Dinont, C., Seghrouchni, A.E.F., Taillibert, P.: A multi-agent approach for the deployment of distributed applications in smart environments. *Intelligent Distributed Computing X*. SCI, vol. 678, pp. 37–46. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-48829-5_4

15. Pokahr, A., Braubach, L., Lamersdorf, W.: Jadex: a BDI reasoning engine. In: Bordini, R.H., Dastani, M., Dix, J., El Fallah Seghrouchni, A. (eds.) *Multi-Agent Programming. MSASSO*, vol. 15, pp. 149–174. Springer, Boston, MA (2005). https://doi.org/10.1007/0-387-26350-0_6
16. Seghrouchni, A.E.F., Florea, A.M., Olaru, A.: Multi-agent systems: a paradigm to design ambient intelligent applications. In: Essaïdi, M., Malgeri, M., Badica, C. (eds.) *Intelligent Distributed Computing IV*, pp. 3–9. Springer, Heidelberg (2010)
17. Twardowski, B., Ryzko, D.: Multi-agent architecture for real-time big data processing. In: *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, vol. 3, pp. 333–337. IEEE (2014)
18. Winikoff, M.: JackTM intelligent agents: an industrial strength platform. In: Bordini, R.H., Dastani, M., Dix, J., El Fallah Seghrouchni, A. (eds.) *Multi-Agent Programming. MSASSO*, vol. 15, pp. 175–193. Springer, Boston, MA (2005). https://doi.org/10.1007/0-387-26350-0_7
19. Zoumpatianos, K., Palpanas, T.: Data series management: Fulfilling the need for big sequence analytics. In: *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pp. 1677–1678. IEEE (2018)



Human Activities Recognition Using Accelerometer and Gyroscope

Anna Ferrari, Daniela Micucci , Marco Mobilio , and Paolo Napoletano  

University of Milano - Bicocca, Milan, Italy
a.ferrari34@campus.unimib.it,
{daniela.micucci,marco.mobilio,paolo.napoletano}@unimib.it

Abstract. Plenty of supervised machine learning techniques that use accelerometer and gyroscope signals for automatic Human Activity Recognition (HAR) has been proposed in the last decade. According to recent studies, the combination of accelerometer and gyroscope signals, also called multimodal recognition, increases the accuracy in HAR with respect to the use of each signal alone. This paper presents the results of an analysis we performed in order to compare the effectiveness of machine learning techniques when used separately or jointly on accelerometer and gyroscope signals. We compare SVM and k -NN classifiers (combined with hand-crafted features) with a deep residual network using three publicly available datasets. The results show that the use of deep learning techniques in multimodal mode (i.e., using accelerometer and gyroscope signals jointly) outperforms other strategies of at least 10%.

Keywords: Inertial sensors · Machine learning · Deep learning · Human Activity Recognition

1 Introduction and Background

The increasingly pervasive connectivity and the growing use of mobile and wearable devices have enabled the diffusion of a new category of applications and services, ranging from smart home automation to remote health-care assistance [1, 8].

Today portable devices have advanced computing capability and connectivity and usually include several sensors, such as a accelerometer and gyroscope, which provide a considerable amount of data. Those factors have stimulated the interest of the scientific community in developing machine learning methods for automatic Human Activity Recognition (HAR). HAR has many applications in several domains such as, for example, healthcare, sport, and entertainment.

During the last decade, a plenty of traditional machine learning as well as deep learning methods for HAR that use inertial sensors has been proposed in literature [13, 16]. Recently, deep learning methods are predominant [16]. These methods require a special hardware setup (Graphical Processing Units - GPUs) to speed up computation and a great amount of data to avoid overfitting during

the training phase. However, it is difficult for low cost consumer hardware to be equipped with GPUs. Thus, in most cases, deep learning methods run on cloud platform, such as, Google Cloud¹, Amazon AWS², or Microsoft Azure³.

So far, no technique has been identified that is more effective than the others. Furthermore, it has not yet been demonstrated that the use of more than one type of signal (i.e., multimodality) provides better results than the use of one type of signal only (i.e., single modality).

Since in HAR large scale inertial datasets are not available [5, 7], it is not obvious which method between deep and traditional machine learning is the most appropriate, especially in the case where hardware is low cost [3]. In addition, it is not yet assessed whether sensorial multimodality (e.g., accelerometer and gyroscope) is more effective than single modality (e.g., accelerometer or gyroscope). The aim of this paper is two-fold: (1) evaluate the effectiveness of hand-crafted features with respect to an end-to-end deep learning based on a Residual Network; (2) evaluate the effectiveness of multimodal HAR with respect to HAR based on accelerometer or gyroscope. Experiments on three public datasets are presented and discussed.

2 Materials and Methods

2.1 Datasets

We experimented three datasets:

- *UCI HAR* [2], which includes tri-axial acceleration and gyroscope data of 6 ADLs (Activities of Daily Living) recorded with a Samsung Galaxy S II and performed by 30 volunteers.
- *MobiAct* [15], which includes tri-axial acceleration, gyroscope, and orientation data of 11 ADLs and 4 Falls recorded with a Samsung Galaxy S3 and performed by 67 volunteers.
- *Motion Sense* [10], which includes tri-axial acceleration and gyroscope data of 6 ADLs recorded with an iPhone 6s and performed by 30 volunteers.

Each signal of the datasets is composed of three accelerometer and three gyroscope components along the x, y, and z axis. Each signal component has been resampled at 50 Hz and divided in segments of 2.56 s with an overlap between subsequent segments of 50% [13]. The resampling at 50 Hz was necessary because the *MobiAct* dataset has been acquired at a frequency of about 87 Hz. The resulting segment for each axis contains 128 samples.

Each dataset has been split in 70% training, 20% validation, and 10% test. Table 1 shows the total number of 128×3 -dimensional segments (128×6 -dimensional in case of multimodal) available for the training (column *# train*), validation (column *# validation*), and test (column *# test*) sets. The last column *# classes* indicates the number of ADLs present in the dataset.

¹ <https://cloud.google.com>.

² <https://aws.amazon.com/>.

³ <https://azure.microsoft.com>.

Table 1. Number of segments and classes for each dataset

Dataset	# train	# validation	# test	# classes
UCI HAR	7,209	2,060	1,030	6
MobiAct	34,070	9,734	4,867	15
Motion Sense	14,945	4,270	2,135	6

2.2 Hand-Crafted Features

For the experimentation of hand-crafted features, the k Nearest Neighbour (k -NN) and Support Vector Machines (SVM) classifiers have been used. The used features are⁴:

- Raw data (denoted as *raw*): x, y , and z of the segments (without any kind of processing) are concatenated and used as feature vectors [12];
- Magnitude of the segments (denoted as *magn*) [11];
- 21 features extracted from the magnitude of the segments (denoted as *hc magn*) [4];
- 21 features extracted from each of the three segments along the three axes x, y , and z (denoted as *hc raw*). The total number of features is 63 [4].

In the case of SVM, the multi-class classifier has been implemented as multiple binary classifiers. We used the Statistics and Machine Learning Toolbox for k -NN and SVM. Optimum parameters of both classifiers have been found through cross-validation using the validation set.

When accelerometer and gyroscope are used in combination, the features are obtained by concatenating the features extracted from each signal.

2.3 End-to-End Deep Learning Solution

The Residual Network (ResNet) adopted for this study is based on a ResNet proposed in [6] and inspired to the traditional architecture proposed by He *et al.* [9], which demonstrated to be very effective on the ILSVRC 2015 (ImageNet Large Scale Visual Recognition Challenge) validation set with a top 1- recognition accuracy of about 80%.

The input size of the network is $1 \times 128 \times 3$, that corresponds to 3 segments along the three axes x, y , and z . In the case of multimodal signals, the input is $1 \times 128 \times 6$. The network architecture is made of an initial convolutional block, 3 residual stages, each containing a variable number n of residual blocks, average pooling layer, fully connected layer, and softmax layer. A convolutional block is made of three layers: convolutional, batch normalization, and ReLu. A residual block is made of 2 subsequent convolutional blocks and an additional operator that sums the input of the residual block with the output of the residual block itself. Each convolutional layer is $1 \times 3 \times f_{maps}$, where f_{maps} is the number of

⁴ We describe the extraction of features from the accelerometer or gyroscope signals.

Table 2. Experimental results - average accuracy (standard deviation): k -NN vs ResNet. For each row, the bold font represents the best.

	Dataset	KNN				ResNet
		raw	magn	hc raw	hc magn	
ACC	UCI-HAR	73.71 (± 26.78)	46.92 (± 29.89)	69.35 (± 17.04)	37.75 (± 13.39)	90.73 (± 10.92)
	MobiAct	87.69 (± 9.07)	77.81 (± 13.60)	91.86 (± 6.72)	80.50 (± 10.74)	92.98 (± 8.65)
	MotionSense	79.19 (± 31.83)	73.51 (± 25.16)	95.82 (± 5.61)	81.34 (± 20.30)	99.47 (± 0.87)
GYRO	UCI-HAR	70.74 (± 24.73)	38.37 (± 25.11)	60.19 (± 20.14)	33.50 (± 6.97)	89.36 (± 9.90)
	MobiAct	78.54 (± 16.13)	73.66 (± 19.38)	83.19 (± 20.43)	74.62 (± 20.44)	96.09 (± 3.16)
	MotionSense	85.16 (± 12.03)	70.75 (± 22.24)	88.74 (± 10.31)	71.69 (± 13.62)	98.07 (± 1.56)
ACC+ GYRO	UCI-HAR	82.36 (± 19.99)	55.35 (± 29.60)	77.10 (± 16.73)	39.74 (± 11.79)	96.46 (± 4.06)
	MobiAct	86.25 (± 8.89)	80.22 (± 14.85)	94.20 (± 5.84)	81.46 (± 11.59)	92.94 (± 9.39)
	MotionSense	74.08 (± 29.86)	73.89 (± 24.70)	97.17 (± 2.33)	84.36 (± 9.45)	99.08 (± 0.65)

Table 3. Experimental results - average accuracy (standard deviation): SVM vs ResNet. For each row, the bold font represents the best.

	Dataset	SVM				ResNet
		raw	magn	hc raw	hc magn	
ACC	UCI-HAR	79.51 (± 17.40)	53.10 (± 25.48)	79.47 (± 20.59)	48.45 (± 22.12)	90.73 (± 10.92)
	MobiAct	77.93 (± 22.71)	63.63 (± 24.13)	76.73 (± 26.11)	59.95 (± 23.94)	92.98 (± 8.65)
	MotionSense	90.04 (± 14.36)	78.22 (± 29.59)	96.39 (± 3.79)	83.45 (± 21.13)	99.47 (± 0.87)
GYRO	UCI-HAR	72.93 (± 23.82)	44.52 (± 27.21)	75.45 (± 14.76)	41.10 (± 17.91)	89.36 (± 9.90)
	MobiAct	64.19 (± 31.37)	57.80 (± 27.49)	68.86 (± 27.70)	52.21 (± 29.22)	96.09 (± 3.16)
	MotionSense	86.92 (± 7.51)	73.93 (± 21.86)	88.32 (± 9.86)	76.46 (± 13.17)	98.07 (± 1.56)
ACC+ GYRO	UCI-HAR	86.83 (± 15.53)	59.49 (± 28.25)	88.14 (± 10.66)	49.20 (± 20.45)	96.46 (± 4.06)
	MobiAct	79.13 (± 18.25)	70.15 (± 22.94)	85.54 (± 16.31)	62.77 (± 23.99)	92.94 (± 9.39)
	MotionSense	85.87 (± 8.05)	80.93 (± 11.55)	95.90 (± 3.07)	85.01 (± 9.29)	99.08 (± 0.65)

feature maps of the filter. For each dataset, the best values for n and f_{maps} have been found by following a grid search approach: n ranged between 3 and 21, while f_{maps} ranged between 10 and 200.

For all the datasets, the networks have been optimized through the Stochastic Gradient Descent with Momentum (SGDM), using a piecewise learning update strategy with an initial value of 0.1 and a drop factor of 0.1. The batch size was 128, the total number of epochs was 80, and the early stopping has been used to avoid overfitting. We used the Matlab Deep Learning Toolbox for training and testing the Residual Network.

3 Experiments

Tables 2 and 3 show results achieved by all the methods considered in terms of macro average accuracy (i.e., the average of each class accuracy). The accuracy of each class is computed as ratio between the number of segments correctly classified and the total number of segments of the given class. ResNet achieves better performance than traditional methods in all datasets apart from MobiAct. This is probably due to the larger number of classes with respect to the other two datasets. Most important, the standard deviation of the ResNet method is close

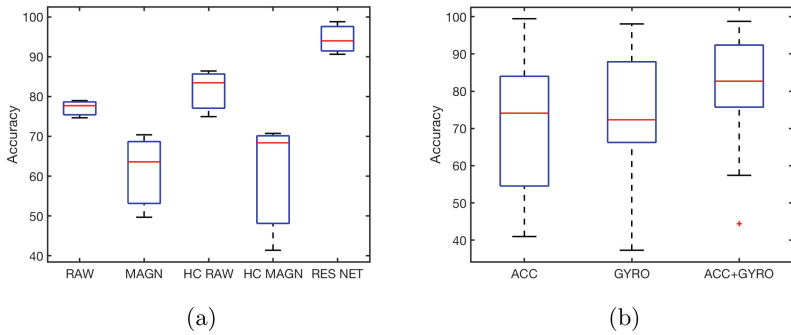


Fig. 1. Experiments. (a) comparison across datasets between hand-crafted and ResNet. (b) comparison across datasets and methods between multimodality and single modality

to zero. Accuracy of k -NN and SVM is quite similar, while, among hand-crafted features, the most performing is the *hc raw*.

Figure 1(a) shows the comparison between ResNet and both k -NN and SVM across dataset and independently of the inertial signal used. Overall, ResNet is the best performing with an average accuracy across datasets of about 93%. Among the traditional classifiers, the best results are achieved by *hc raw* features followed by the *raw* ones. The worst results are obtained by using magnitudes (both *hc magnitude* and *magnitude raw*).

Figure 1(b) shows the comparison across datasets, between methods based on sensorial multimodality (i.e., accelerometer and gyroscope - ACC+GYRO) and single modality (i.e., accelerometer or gyroscope - ACC or GYRO). Overall, multimodal recognition works better than the single modal with an improvement of about 10%. Accelerometer is more performing than the gyroscope. This result is confirmed by the fact that most of the experiments undertaken in the literature are based on accelerometric signals only [14].

In summary, the average gap between hand-crafted features combined with traditional classifiers and deep learning is about 10% thus confirming that, on these datasets, deep learning approaches outperforms traditional ones.

4 Conclusions

This paper presents a comparison between traditional classifiers combined with hand-crafted features and an end-to-end deep learning solution based on a Residual Network. Experiments on three public datasets demonstrated that overall deep learning solutions overcome traditional machine learning approaches and that, the joint use of accelerometer and gyroscope allows to increase performance of about 10% with respect to the use of accelerometer or gyroscope alone.

As future work, we will evaluate the robustness of deep learning techniques across position of the wearable device, age, height, and gender of the subject.

References

1. Acampora, G., Cook, D., Rashidi, P., Vasilakos, A.: A survey on ambient intelligence in healthcare. *Proc. IEEE* **101**(12), 2470–2494 (2013)
2. Anguita, D., Ghio, A., Oneto, L., Parra, X., Reyes-Ortiz, J.L.: A public domain dataset for human activity recognition using smartphones. In: *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN13)* (2013)
3. Bianco, S., Cadene, R., Celona, L., Napoletano, P.: Benchmark analysis of representative deep neural network architectures. *IEEE Access* **6**, 64270–64277 (2018)
4. Bianco, S., Napoletano, P., Schettini, R.: Multimodal car driver stress recognition. In: *Proceedings of the EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth19)* (2019)
5. Ferrari, A., Micucci, D., Marco, M., Napoletano, P.: A framework for long-term data collection to support automatic human activity recognition. In: *Proceedings of the Intelligent Environments (IE) Work on Reliable Intelligent Environment (WoRIE19)* (2019)
6. Ferrari, A., Micucci, D., Marco, M., Napoletano, P.: Hand-crafted features vs residual networks for human activities recognition using accelerometer. In: *Proceedings of the IEEE International Symposium on Consumer Technologies (ISCT19)* (2019)
7. Ferrari, A., Micucci, D., Marco, M., Napoletano, P.: On the homogenization of heterogeneous inertial-based databases for human activity recognition. In: *Proceedings of IEEE Services Work on Big Data for Public Health Policy Making* (2019)
8. Grossi, G., Lanzarotti, R., Napoletano, P., Noceti, N., Odone, F.: Positive technology for elderly well-being: a review. *Pattern Recogn. Lett.* (2019)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR16)* (2016)
10. Malekzadeh, M., Clegg, R.G., Cavallaro, A., Haddadi, H.: Protecting sensory data against sensitive inferences. In: *Proceedings of the Workshop on Privacy by Design in Distributed Systems (W-P2DS18)* (2018)
11. Micucci, D., Mobilio, M., Napoletano, P.: Unimib shar: a dataset for human activity recognition using acceleration data from smartphones. *Appl. Sci.* **7**(10), 1101 (2017)
12. Micucci, D., Mobilio, M., Napoletano, P., Tisato, F.: Falls as anomalies? an experimental evaluation using smartphone accelerometer data. *J. Ambient Intell. Humaniz. Comput.* **8**(1), 87–99 (2017)
13. Ronao, C.A., Cho, S.B.: Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Syst. Appl.* **59**, 235–244 (2016)
14. Twomey, N., et al.: Comprehensive study of activity recognition using accelerometers. *Informatics* **5**(2), 27 (2018)
15. Vavoulas, G., Chatzaki, C., Malliotakis, T., Padiaditis, M., Tsiknakis, M.: The mobiact dataset: recognition of activities of daily living using smartphones. In: *Proceedings of Information and Communication Technologies for Ageing Well and e-Health (ICT4AgeingWell16)* (2016)
16. Wang, J., Chen, Y., Hao, S., Peng, X., Hu, L.: Deep learning for sensor-based activity recognition: a survey. *Pattern Recogn. Lett.* **119**, 3–11 (2019)



Towards Habit Recognition in Smart Homes for People with Dementia

Gibson Chimamiwa^(✉), Marjan Alirezaie, Hadi Banaee, Uwe Köckemann,
and Amy Loutfi

Centre for Applied Autonomous Sensor Systems (AASS), Örebro, Sweden
{gibson.chimamiwa,marjan.alirezaie,hadi.banaee,
uwe.kockemann,amy.loutfi}@oru.se

Abstract. The demand for smart home technologies that enable ageing in place is rising. Through activity recognition, users' activities can be monitored. However, for dementia patients, activity recognition alone cannot address the challenges associated with changes in the user's habits along the disease's stage transitions. Extending activity recognition to habit recognition enables the capturing of patients' habits and changes in habits in order to detect anomalies. This paper aims to introduce relevant features for habit recognition solutions, extracted from data, in order to enrich the representation of the user's habits. This solution is personalisable to meet the specific needs of the patients and generalizable for use in different scenarios. In this way caregivers are better informed on the expected changes of the patient's habits, which can help to mitigate further deterioration through early treatment and intervention.

Keywords: Habit recognition · Dementia · Smart homes

1 Introduction

Due to the growing number of people with dementia and the high societal cost of the disease [7], the demand for the development of smart homes to support patients is increasing [3]. Dementia progresses in seven developmental stages [9]. A major challenge in dementia is early recognition and perhaps prevention of dementia stage transition of patients [14]. Due to cognitive decline, patients are prone to change their habits at different stages of the disease. A habit is one or a set of activities that an individual usually does in a regular and repeated way [12]. An example of a habit is *usually resting after having dinner*. Detecting such habits can assist to determine stage transition in the course of dementia.

The state of the art in smart home systems for health care purposes relies on Activity Recognition (AR) to monitor daily activities of users [11]. To monitor people with dementia and deal with the challenge of dementia stage transition, solutions based on smart home technologies need to go beyond activity recognition and be able to recognise the habits of the patient. Habit recognition can be a solution to capture the current stage of the disease, which enables the detection

of changes in the habits in order to detect anomalies. In addition, each individual patient experiences the changes in a unique way, thus dementia is highly personalised. Therefore, the solution for habit recognition should be personalisable to assist specific patients and generalizable to be re-usable for different scenarios.

Most AR solutions in the literature for dementia have not given much attention to human habits, which is critical for assisting dementia patients. Solutions for activity recognition in smart homes are categorized into data-driven and knowledge-driven approaches [8]. Knowledge-driven approaches depend on the availability of domain experts to create activity models based on prior knowledge about the environment. The advantage of such approaches is that the situations of interest (e.g., activities and events) can be modeled in advance and be recognized by an automated reasoner [2, 10, 13]. Data-driven approaches, on the other hand, are focused on extracting activity patterns from raw sensor data, for instance detecting anomalous behaviour of dementia patients using wearable sensors [11] or recognizing specific behaviours of dementia patients such as wandering based on repeated indoor movements detected by environmental sensors [6]. Through activity recognition, anomalous behaviours of dementia patients such as repeating activities, sleep disruption, and confusion are also detected [4]. Furthermore, a pattern recognition method has been developed using medical sensor data to detect changes in activity patterns for early signs of cognitive and health decline [5]. Although data-driven methods have performed well in activity recognition, the main drawback of such methods is the availability of (good) data representing cognitive decline that may be difficult to acquire.

In this paper, we propose our habit recognition solution which extracts features from data to enrich the representation of the user's habits. This solution will be used to extend the available knowledge-driven activity recognition system, E-care@home [2]. In Sect. 2, we briefly present the E-care@home system as a knowledge-driven solution for AR. We then discuss our habit recognition approach in Sect. 3 as an extension to the E-care@home system. In Sect. 4, we state potential benefits of the solution for dementia patients and present ideas for future work.

2 E-care@home System

The context recognition system in E-care@home is a knowledge-driven process. As shown in Fig. 1, the knowledge model is an ontology called SmartEnv (Smart Environment ontology) [1]. SmartEnv provides the representational basis required to model an environment equipped with various sensors including spatio-temporal facets of the environment. The sensors include pressure, motion, and luminosity to detect activities resting, eating, being in bathroom, and cooking [2]. In SmartEnv, activities are defined as sub-classes of the class *Event*, which are categorized into two classes of *Manifestation* and *ComplexEvent*. A manifestation refers to an activity or a change in a given environment that can be recognized directly from sensor data (e.g., `manifestation(couch, true, pressed, t1)`, to indicate that the *pressure* sensor under the *couch* is triggered at t_1).

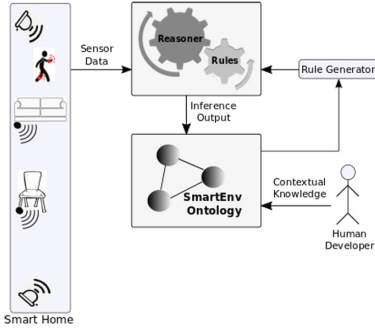


Fig. 1. E-care@home system architecture.

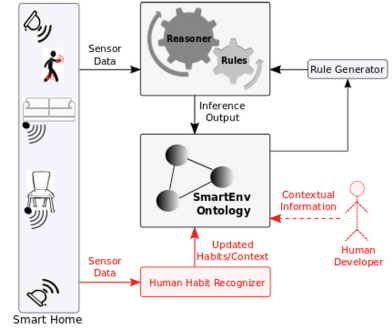


Fig. 2. Extended system with habit recognition. (Color figure online)

Although a manifestation can reflect an activity (e.g., sitting on the couch) there are other situations that are a combination of several manifestations (i.e., pre-conditions). The class *ComplexEvent* is defined to model such situations. Each complex event is defined based on its preconditions.

The context recognition in the E-care@home system starts by reading sensor data and transferring it to the reasoner. Given the data, the reasoner which is already initialized with logic programs generated from the ontology, results in time intervals labeled with activities' names, events or other information inferred by the reasoner. The logic programs contain logic rules that represent events in terms of their preconditions (defined in the ontology) as follows:

$$event(E, t) \leftarrow eventCondition(C_1, t), \dots, eventCondition(C_n, t). \quad (1)$$

Each *eventCondition* predicate represents a manifestation and its temporal relations (T_1, T_2) w.r.t to a timestamp at which the event is expected to occur.

$$eventCondition(C_1, t) :- manifestation(O_1, P_1, S_1, T_1, T_2), T_1 \leq T_2 \quad (2)$$

For instance, the activity *eating* is defined based on 3 preconditions of *having recently cooked* (e.g., 900s ago), *being in the kitchen* and *sitting on the chair*:

$$event(eating, t) \leftarrow eventCondition(ec1, t), eventCondition(ec2, t), eventCondition(ec3, t).$$

$$where: eventCondition(ec1, t) \leftarrow event(cooking, t - 900).$$

$$eventCondition(ec2, t) \leftarrow manifestation(kitchen, motion, true, t).$$

$$eventCondition(ec3, t) \leftarrow manifestation(chair, pressure, true, t).$$

Although *ComplexEvent* allows us to define activities based on preconditions and their temporal relations, it is not practical to manually model all the details reflecting how an individual performs an activity. In other words, manual representation of human habits is an intractable process. The human habit recognizer shown in red in Fig. 2 is a data-driven module able to capture habits of the user in doing an individual activity or as correlations between activities.

3 Human Habit Recognition

As shown in Fig. 2, to capture habits, the E-care@home system will be extended with a data-driven module that extracts several features of data reflecting how and when the user is involved in different daily activities. The data used for the evaluation of the implemented module was gathered constantly during 10 days where the user also labeled some of the activities including *resting*, *sitting*, *cooking*, *eating*, *watching_tv*, and *being_in_bathroom* in order to provide a comparative ground truth. The data is in the form of labeled intervals recorded in the database of E-care@home which is publicly available at: <https://ecaredb.oru.se> (user name and password: *ecare-pub*).

The features that reflect habits of the user in doing daily activities can include: the duration or the frequency of performing an activity (e.g., *the user usually sleeps for 5 h per night*), regular interruptions (e.g., *the user usually wakes up 3 times per night*) or regular absence of doing an activity (e.g., *the user does not usually rest after eating*) as well as co-occurrence of multiple activities (e.g., *the user is usually watching TV while eating*). At this phase of the work, we only focus on the duration and frequency for the individual activities explained in the following subsections.

Three methods have been implemented to capture the duration and frequency features. These features can later be used to recognize important changes in the habits of the user to show signs of cognitive decline. We use the following notations to clarify details in the implementation:

- *dur* is the duration of each activity calculated as *end time - start time*.
- *freq* represents the number of times an activity occurs in a given time period.
- *n* represents the number of days within the monitoring period.
- $t_{md}, t_{em}, t_{mn}, t_{nn}, t_{an}, t_{ev}, t_{ni}$ represent the times of a day, namely *midnight*, *early morning*, *morning*, *noon*, *afternoon*, *evening*, and *night* respectively.

Method 1: Average frequency per day is the average of the frequency of an activity in each day. For instance, if the frequency of the activity *eating* in day *i* is $freq_{eat,i}$, then the average frequency of *eating* is calculated as: $\overline{freq}_{eat} = (\sum_{i=1}^n freq_{eat,i})/n$.

Calculating the average frequency can be beneficial to detect the changes in how often an activity occurs. For example, a noticeable increase in the frequency of *eating* activity may signal a problem of forgetfulness where dementia patient may eat repeatedly. From the results shown in Table 1, we observe that, on average the user rests, cooks and eats 2 times per day and uses the bathroom 9 times per day.

Method 2: Average duration is the average of activity duration within the monitoring period. For instance, the average duration for activity *resting* in day *i* is calculated as: $\overline{dur}_{rest,i} = (\sum_{j=1}^{freq_{rest,i}} dur_{rest,i,j})/freq_{rest,i}$.

Given the daily average of the durations, the overall average of the durations for the activity will be: $\overline{dur}_{rest} = (\sum_{i=1}^n \overline{dur}_{rest,i})/n$.

Computing the average duration can be beneficial to detect the changes of how long an activity occurs. For instance, an increase in the average duration spent on *resting* may indicate the later stages of dementia associated with increased body weariness. Table 1 shows results of the average duration for the 4 activities *resting*, *cooking*, *eating* and *being.in.bathroom* (e.g., *resting* lasts for 5 h 54 m).

Method 3: Average frequency per time of day is the average of an activity frequency in each labeled time of day (e.g., *noon*, *afternoon*, etc.). For instance, if $freq_{eat_{t_{nn}},i}$ indicates the frequency of activity *eating* at *noon* for day *i*, then the overall average of the frequency of activity *eating* at *noon* is calculated as: $(\sum_{n=1}^n freq_{eat_{t_{nn}},i})/n$.

Likewise, calculating the average frequency per time of day can help to determine if the user performs activities at regular intervals. For instance, a reduction in the average frequency of activity *eating* at certain time of the day (e.g., *noon*) may indicate that the patient is forgetting to have lunch. From the results shown in Table 2, the user usually rests once between midnight and morning. Cooking and eating usually happen once at noon and at night, and the user does not use the bathroom during early morning. By including the features captured directly from the data into the ontology, we can personalize the knowledge model for a specific user. In this way, we can deal with the problem of manually modeling all the details reflecting the habits of an individual when performing an activity. Each of these captured features can be modeled as a new precondition that forms an axiom defining an activity in the ontology (see Eq. 1). Furthermore, the reasoner fed with more personalized data, can better recognize patterns of activities as well as sensitive changes in the recognized habits.

Table 1. Average frequency per day and average duration of activities

	resting	cooking	eating	in_bathroom
Average frequency per day	2	2	2	9
Average duration	5 h 54 m	16 m 18 s	18 m 25 s	5 m 3 s

Table 2. Average frequency of habit per time of day

	Midnight	Early morning	Morning	Noon	Afternoon	Evening	Night
resting	1	1	1	–	–	–	–
in_bathroom	1	–	2	1	1	2	2
eating	–	–	–	1	–	–	1
cooking	–	–	–	1	–	–	1

4 Conclusions and Future Work

In this paper we have presented our ongoing work towards habit recognition for people with dementia. The proposed solution assists the process of detecting changes in the habits of dementia patients. In this way, caregivers and patients can be made aware of what to expect in the following phase of the disease as well as the ability to initiate early treatment of the disease. This work is at its early stage of development and will be extended to include more general and autonomous feature extraction methods. These methods will be able to capture more complex habits including correlations between activities (e.g., *the user usually eats while watching TV*) or absence of doing an activity (e.g., *the user does not usually rest after eating*) that might be of interest in predicting cognitive decline well in advance and detecting dementia stage transition.

Acknowledgments. This work has been supported by both the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754285, and the distributed research environment E-care@home funded by the Swedish Knowledge Foundation (KKS), 2015–2019.



References

- Alirezaie, M., Hammar, K., Blomqvist, E.: Smartenv as a network of ontology patterns. *Semant. Web* **9**(6), 903–918 (2018)
- Alirezaie, M., et al.: An ontology-based context-aware system for smart homes: E-care@ home. *Sensors* **17**(7), 1586 (2017)
- Amiribesheli, M., Bouchachia, H.: A tailored smart home for dementia care. *J. Ambient Intell. Hum. Comput.* **9**(6), 1755–1782 (2018)
- Arifoglu, D., Bouchachia, A.: Detection of abnormal behaviour for dementia sufferers using convolutional neural networks. *A.I. Med.* **94**, 88–95 (2019)
- Enshaeifar, S., et al.: Machine learning methods for detecting urinary tract infection and analysing daily living activities in people with dementia. *PLoS ONE* **14**(1), e0209909 (2019)
- Lin, Q., Zhao, W., Wang, W.: Detecting dementia-related wandering locomotion of elders by leveraging active infrared sensors. *J. Comput. Commun.* **6**(05), 94 (2018)
- Prince, M., Wimo, A., Guerchet, M., Ali, G., Wu, Y., Prina, M.: World Alzheimer report 2015—the global impact of dementia, an analysis of prevalence, incidence, cost and trends. *Alzheimer's Dis. Int.* **17**, 2016 (2015)
- Ranasinghe, S., Al Machot, F., Mayr, H.C.: A review on applications of activity recognition systems with regard to performance and evaluation. *Int. J. Distrib. Sens. Netw.* **12**(8) (2016). <https://doi.org/10.1177/1550147716665520>
- Reisberg, B., Ferris, S.H., de Leon, M.J., Crook, T.: The global deterioration scale for assessment of primary degenerative dementia. *Psychiatry J.* **139**(9), 1136–1139 (1982)
- Stavropoulos, T.G., Meditskos, G., Andreadis, S., Avgerinakis, K., Adam, K., Kompatsiaris, I.: Semantic event fusion of computer vision and ambient sensor data for activity recognition to support dementia care. *J. Ambient. Intell. Hum. Comput.* 1–16 (2016). <https://doi.org/10.1007/s12652-016-0437-5>

11. Su, C.F., Fu, L.C., Chien, Y.W., Li, T.Y.: Activity recognition system for dementia in smart homes based on wearable sensor data. In: 2018 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 463–469. IEEE (2018)
12. Thompson, M.: Occupations, habits, and routines: perspectives from persons with diabetes. *Scand. J. Occup. Ther.* **21**(2), 153–160 (2014)
13. Tiberghien, T., Mokhtari, M., Aloulou, H., Biswas, J.: Semantic reasoning in context-aware assistive environments to support ageing with dementia. In: Cudré-Mauroux, P., et al. (eds.) ISWC 2012. LNCS, vol. 7650, pp. 212–227. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-35173-0_14
14. Woods, R.T., et al.: Dementia: issues in early recognition and intervention in primary care. *J. R. Soc. Med.* **96**, 320–324 (2003)



Ambient Explanations: Ambient Intelligence and Explainable AI

Jörg Cassens¹(✉)  and Rebekah Wegener² 

¹ University of Hildesheim, 31141 Hildesheim, Germany
cassens@cs.uni-hildesheim.de

² Paris Lodron University of Salzburg, 5020 Salzburg, Austria
rebekah.wegener@sbg.ac.at

Abstract. With renewed prominence of Explainable AI (XAI), many areas are revisiting early work on Explainability. Within the broader field of Artificial Intelligence (AI), Ambient Intelligence (AmI) has an advantage in the development of transparent and ethical systems because such work has long been an integral part of research, development and operations in AmI. In this paper we argue that, because of the paradigm requirements of system intelligence, social intelligence and embeddedness, AmI is uniquely prepared to support the push for ethical and transparent technology development. We argue that Ambient Intelligent Systems are well suited to infer when an explanation might be needed (and of what kind), and the form that it should take. We further propose AmI devices as mediators between humans and machines because they can combine social and technical systems in a fully embedded way.

Keywords: Explanations · Semiotics · Context · Ambient Intelligence

1 Introduction

Explanations are crucial to understanding, cognition, reasoning, discovery and learning as well as developing a sense of self [20]. They are also the foundation of social interaction [19]. While there is compelling research supporting the value, structure and function of explanation, “accounts of explanation typically define explanation (the product) rather than explaining (the process)” [7]. We have earlier construed contextualised explanations based on user goals [27]. This has been used to integrate explanatory needs in the system design process [3, 22]. However, we have represented explanation as a static object rather than a dialogic process. In our current work, we look at the process of explanatory behaviour instead, including what triggers explanatory behaviour in different contexts and how explanations might be evaluated. More importantly, we look at the dialogic characteristics of explanatory behaviour and the impact of contextual variation.

It has been shown that context makes a difference not only to the type of explanation used, but also to how explanations are evaluated [28]. The authors found that “people evaluate the quality of explanations differently depending on

how well the explanations suite their needs in a given context”, although this is on top of a clear preference for certain types. They also suggest that further work is needed to examine the scope and basis of the context effect. To test this effect, we propose to make use of the semiotic parameters of field, tenor and mode [12, 13]. Semiotic models of context have been used for ambient intelligence before [4, 29], but they have to be extended to include explanatory behaviour.

Explanations are a long established research topic in computer science and artificial intelligence [16, 23, 24, 26, 27]. They are interesting both as a means of reasoning and for human-computer interaction. The human and social sciences provide a rich basis for applications in explainable AI [21]. Increased use of principles and methods from Artificial Intelligence (AI) and Machine Learning (ML) in applications domains has recently led to new research programs focusing on fairness, interpretability and explainability in ML [18], on accountability, responsibility and transparency in AI [8] and to a renewed interest in explainable AI (XAI) [10].

Our working definition for explanation awareness is as follows:

- **Internal View:** Explanation as *part of the reasoning process* itself.
 - *Example:* a recommender system can use domain knowledge to explain the absence or variation of feature values, e.g. relations between countries
- **External View:** giving explanations of the found solution, its application, or the reasoning process *to the other actors*
 - *Example:* the user tells said recommender system why he chooses an apartment in Norway despite the system suggesting one in Sweden

However, explanations are currently largely seen as a relatively uniform and definable concept, and even systems that take user goals with explanation into account treat it largely on the system side of development [2]. By contrast, there is sufficient empirical and theoretical evidence that explanations are generated, communicated, understood and used in ways that are:

- **Dialogic**, as suggested e.g. by Leake [17],
- **Contextualised**, as required by e.g. Fraassen [9], comprised of
 - *Context Awareness* (knowing the situation the system is in) and
 - *Context Sensitivity* (acting according to such situation) [14, 15]
- **Multimodal**, as argued for by e.g. Halliday [11] and being
- **Construed by user interest**, as noted by e.g. Achinstein [1].

Given these foundations, can a semiotic model of explanation as a form of multi-modal dialogic language behaviour in context be used to generate contextually appropriate explanations by computational systems? There is an extensive body of research focusing on generating and using explanations in AI. Currently, what is lacking is:

1. A theory of the **dialogic process** rather than a monologic product
2. A **cohesive theory of explanation** that is:
 - *contextually appropriate* (e.g. fitting people, topic, mode and place),
 - *semantically appropriate* (e.g. recognised as an explanation)
 - *lexicographically optimal* (best possible multi-modal realisation)

3. A framework for integrating explanatory capabilities in the whole **software development life-cycle**, from requirements elicitation over design and implementation through to its use
4. A framework for **evaluation measures** that is:
 - *intrinsic* (deciding on a strategy for explanation generation)
 - *dialogic* (measuring the reaction to an explanation and providing further explanation if needed)

We suggest that Ambient Intelligence is intrinsically well suited to achieve the goals of a research agenda targeting these shortcomings.

2 Explanations and Ambient Intelligence

Our working definition for Ambient Intelligence in the context of this paper builds on de Ruyter and Aarts who divide the characteristics of AmI systems into system intelligence, social intelligence and embeddedness [5]. Within their model, AmI is characterized by systems and technologies that are:

- **Embedded:** many networked devices are integrated into the environment;
- **Context aware:** devices can recognize you and your situational context;
- **Personalized:** they can be tailored to your needs;
- **Adaptive:** they can change in response to you;
- **Anticipatory:** anticipating your desires without conscious mediation.

We propose that the area of Ambient Intelligence (AmI) provides the ideal development ground for this work for three key reasons:

1. the core characteristics of AmI systems provide the necessary framework for good explanatory systems,
2. on an abstract level, the architectures for AmI systems are the same as for good explanatory systems, and
3. AmI systems themselves have the potential to become explanatory agents that can be mediators between humans and other systems.

Core Characteristics: Firstly, we argue that the characteristics of AmI as outlined above (embeddedness, context awareness, personalization, adaptiveness, and anticipatoriness) mean that ambient systems should be well suited to produce good explanations. We contend that said requirements for Ambient Intelligence can be translated to explanatory systems as follows:

- Multimodal (Embedded): explanations are generated when and where they are needed, using the best available modalities;
- Context aware (context aware): explanations are aware of specific persons and their situational context;
- Context sensitive (anticipatory): they can anticipate explanatory needs without conscious mediation.

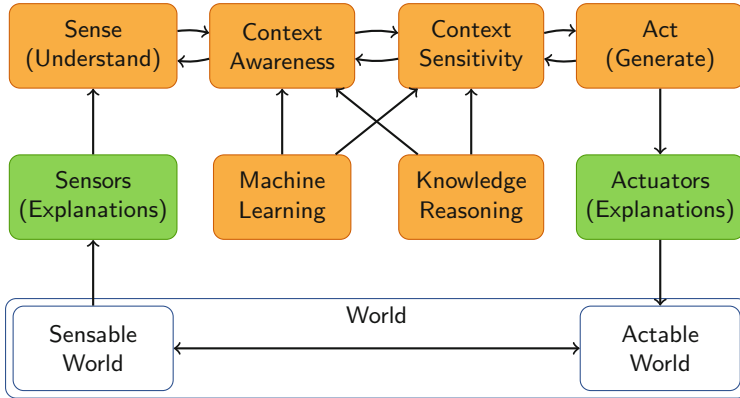


Fig. 1. Generalized abstract architecture for ambient intelligent (explanatory) systems.

- Construed by user interests (Personalized): explanations can be tailored to a person’s needs;
- Dialogic (adaptive): they can change in response to a person’s actions;

Architectures: Secondly, and perhaps most importantly, the abstract architecture for ambient intelligent systems of sensors, processors and actuators is the same as that needed for good explanatory systems.

In Fig. 1, we outline a generalized abstract architecture for ambient intelligent systems, building upon and extending the three layer architecture separating *sensing*, *awareness* (recognising the situation the system is in) and *sensitivity* (acting according to the situation) [14, 15]. Our addition of *acting* completes the feedback loop with the surrounding world.

The figure also shows an instantiation of this general abstract architecture for explanatory systems. Most of the current discussion around XAI is reduced to a system’s ability to generate explanations about its internal state, reasoning process or decisions. In contrast to this, the proposed general architecture also caters for the ability of the system to incorporate explanations given by the user. This could, for example, help identifying new and unseen contexts. We would argue that the ability to integrate new situation is a prerequisite for a model of context that moves from a parametric view to an activity-centred view [6]. In addition to the external view on (user-facing) explanations, integration of internal explanation (i.e. as part of the reasoning process) is also catered for.

Ambient Mediators: Finally, extending the previous two points, it should be possible to build Autonomous Explanatory Agents. Because of the characteristics of AmI and the abstract architecture of networked AmI systems, ambient intelligent systems themselves may make good explainers either as an ambient explanatory space or an autonomous explanatory agent. It is possible that

ambient intelligent systems may be well suited to mediating between humans and other AI systems, or indeed other computational systems in general.

Recently, we have seen a swing towards smart data – or small, richly annotated and contextualised data sets, a re-evaluation of machine learning techniques, and a move towards ethical AI. If we follow the assessment by Streitz et al. [25] that “ethical design is not news to the ambient intelligence community” and that “ethical concerns are inscribed into the conception of ambient intelligence research ab initio”, this movement should be a boon for AmI. Unfortunately, the authors also contend that the development of ethical guidelines is “taking place largely outside of AmI” [25].

3 Conclusion

We have argued that research in Ambient Intelligence has paved the way for further work on explainable systems in a number of different ways. Firstly, core characteristics of ambient systems point towards ambient systems having the necessary characteristics to produce good explanations, that is explanations that are dialogic, contextualised, context aware and context sensitive, multimodal and construed by user interest. These characteristics are not necessarily present in other systems. Secondly, an abstract system architecture of (networked) ambient intelligent systems with sensors, processing units and actuators is similar to what is needed for good explanations. Finally, we argue that, perhaps because of these two aspects (characteristics and system structure) and even with the constraints, ambient intelligent systems themselves may make good explainers either as an explanatory space or an ambient explanatory agent.

References

1. Achinstein, P.: *The Nature of Explanation*. Oxford University Press, Oxford (1983)
2. Biran, O., Cotton, C.: Explanation and justification in machine learning: a survey. In: *IJCAI 2017 Workshop on Explainable AI (XAI)* (2017)
3. Cassens, J., Kofod-Petersen, A.: Explanations and case-based reasoning in ambient intelligent systems. In: Wilson, D.C., Khemani, D. (eds.) *ICCBR 2007 Workshop Proceedings*, Belfast, Northern Ireland, pp. 167–176 (2007)
4. Cassens, J., Wegener, R.: Making use of abstract concepts—systemic-functional linguistics and ambient intelligence. In: Bramer, M. (ed.) *IFIP AI 2008. ITIFIP*, vol. 276, pp. 205–214. Springer, Boston, MA (2008). https://doi.org/10.1007/978-0-387-09695-7_20
5. De Ruyter, B., Aarts, E.: Experience research: a methodology for developing human-centered interfaces. In: Nakashima, H., Aghajan, H., Augusto, J.C. (eds.) *Handbook of Ambient Intelligence and Smart Environments*, pp. 1039–1067. Springer, Boston (2010). https://doi.org/10.1007/978-0-387-93808-0_39
6. Dourish, P.: What we talk about when we talk about context. *Pers. Ubiquit. Comput.* **8**(1), 19–30 (2004)
7. Edwards, B.J., Williams, J.J., Gentner, D., Lombrozo, T.: Explanation recruits comparison in a category-learning task. *Cognition* **185**, 21–38 (2019)

8. Floridi, L., et al.: AI4people – an ethical framework for a good ai society: opportunities, risks, principles, and recommendations. *Mind. Mach.* **28**(4), 689–707 (2018)
9. van Fraassen, B.C.: *The Scientific Image*. Clarendon Press, Oxford (1980)
10. Gunning, D., Aha, D.: Darpa’s explainable artificial intelligence (XAI) program. *AI Mag.* **40**(2), 44–58 (2019)
11. Halliday, M.A.: *Language as a Social Semiotic: The Social Interpretation of Language and Meaning*. University Park Press, Baltimore (1978)
12. Hasan, R.: Situation and the definition of genre. In: Grimshaw, A. (ed.) *What’s Going on Here? Complementary Analysis of Professional Talk: Volume 2 of the Multiple Analysis Project*. Ablex, Norwood (1994)
13. Hasan, R.: Speaking with reference to context. In: Ghadessy, M. (ed.) *Text and Context in Functional Linguistics*. John Benjamins, Amsterdam (1999)
14. Kofod-Petersen, A., Aamodt, A.: Contextualised ambient intelligence through case-based reasoning. In: Roth-Berghofer, T.R., Göker, M.H., Güvenir, H.A. (eds.) *ECCBR 2006. LNCS (LNAI)*, vol. 4106, pp. 211–225. Springer, Heidelberg (2006). https://doi.org/10.1007/11805816_17
15. Kofod-Petersen, A., Cassens, J.: Modelling with problem frames: explanations and context in ambient intelligent systems. In: Beigl, M., Christiansen, H., Roth-Berghofer, T.R., Kofod-Petersen, A., Coventry, K.R., Schmidtke, H.R. (eds.) *CONTEXT 2011. LNCS (LNAI)*, vol. 6967, pp. 145–158. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24279-3_17
16. Leake, D.B.: *Evaluating Explanations: A Content Theory*. Lawrence Erlbaum Associates, New York (1992)
17. Leake, D.B.: Goal-based explanation evaluation. In: *Goal-Driven Learning*, pp. 251–285. MIT Press, Cambridge (1995)
18. Lisboa, P.J.G.: Interpretability in machine learning – principles and practice. In: Masulli, F., Pasi, G., Yager, R. (eds.) *WILF 2013. LNCS (LNAI)*, vol. 8256, pp. 15–21. Springer, Cham (2013). https://doi.org/10.1007/978-3-319-03200-9_2
19. Lombrozo, T.: The structure and function of explanations. *Trends Cogn. Sci.* **10**(10), 464–470 (2006)
20. Lombrozo, T.: The instrumental value of explanations. *Philos. Compass* **6**(8), 539–551 (2011)
21. Miller, T.: Explanation in artificial intelligence: insights from the social sciences. *Artif. Intell.* (2018)
22. Roth-Berghofer, T.R., Cassens, J.: Mapping goals and kinds of explanations to the knowledge containers of case-based reasoning systems. In: Muñoz-Ávila, H., Ricci, F. (eds.) *ICCBR 2005. LNCS (LNAI)*, vol. 3620, pp. 451–464. Springer, Heidelberg (2005). https://doi.org/10.1007/11536406_35
23. Schank, R.C.: *Explanation Patterns - Understanding Mechanically and Creatively*. Lawrence Erlbaum, New York (1986)
24. Shortliffe, E.H.: *Computer-based Medical Consultations: MYCIN*, New York (1976)
25. Streitz, N., Charitos, D., Kaptein, M., Böhlen, M.: Grand challenges for ambient intelligence and implications for design contexts and smart societies. *J. Ambient. Intell. Smart Environ.* **11**(1), 87–107 (2019)
26. Swartout, W.R.: What kind of expert should a system be? XPLAIN: a system for creating and explaining expert consulting programs. *Artif. Intell.* **21**, 285–325 (1983)
27. Sørmo, F., Cassens, J., Aamodt, A.: Explanation in case-based reasoning - perspectives and goals. *Artif. Intell. Rev.* **24**(2), 109–143 (2005)

28. Vasilyeva, N., Wilkenfeld, D., Lombrozo, T.: Contextual utility affects the perceived quality of explanations. *Psychon. Bull. Rev.* **24**(5), 1436–1450 (2017)
29. Wegener, R., Cassens, J., Butt, D.: Start making sense: systemic functional linguistics and ambient intelligence. *Rev. d'Intelligence Artif.* **22**(5), 629–645 (2008). <https://doi.org/10.3166/ria.22.629-645>

Author Index

- Akrivopoulos, Orestis 344
Albayrak, Sahin 203
Alhersh, Taha 123
Alirezaie, Marjan 363
Alves, Ana 296
Amaxilatis, Dimitrios 186, 344
Anagnostopoulos, Aris 235
André, Elisabeth 266, 324
Arsenopoulou, Natalia 283
Aslan, Ilhan 266, 324
- Bacci, Federico 235
Banaee, Hadi 363
Bao, Yuanyuan 92
Bitzas, Dimitris 344
Bourg, Lorena 108
Brahim Belhaouari, Samir 123
- Cassens, Jörg 370
Castelluccio, Carlotta 337
Charitos, Dimitris 283
Chatzidimitris, Thomas 108
Chatzigiannakis, Ioannis 108, 235, 344
Chen, Wai 92
Chimamiwa, Gibson 363
Concas, Emanuele 311
Cruz, Patricio 170
Cuffaro, Giovanni 75
Cunha, Inês 296
Cuomo, Francesca 58
- Dang, Chi Tai 266, 324
De Ruyter, Boris 1
Delinikolas, Dimitris 283
Dessi, Danilo 311
Dietz, Michael 324
Drymonitis, Alexandros 283
- Eryilmaz, Elif 203
- Falcone, Giacomo 337
Farasin, Alessandro 337
Ferrari, Anna 357
- Filipenko, Michael 324
Frengkou, Elton 219
Fröhlich, Peter 41
- Garlisi, Domenico 58
Garofalakis, John 219
Gavalas, Damianos 108
Gerina, Federica 318
Giannakopoulou, Kalliopi 108
Gilgien, Matthias Felix 330
Gomes, Rui 296
- Iannucci, Enrico 350
- Kallio, Johanna 27
Karkazis, Panagiotis 235
Kasapakis, Vlasios 108
Khan, Manzoor Ahmed 203
Khan, Vassilis-Javed 139
Köckemann, Uwe 363
Koivusaari, Jani 27
Kominos, Andreas 219
Konstantopoulos, Charalampos 108
Korosidis, Antonios 283
Kurnyts'kyi, Taras 14
Kyllönen, Vesa 27
Kypriadis, Damianos 108
- Lalos, Aris S. 344
Lin, Fuchun Joseph 153
Liu Cheng, Alexander 170
Llorca Vega, Nestor 170
Loutfi, Amy 363
- Maiano, Luca 235
Makara, Dmytro 14
Martino, Alessio 58
Mastrorilli, Federico 350
Mavrommati, Irene 139
Medrano, Johan 153
Mena, Andrés 170
Micucci, Daniela 357
Mittal, Govind 252

- Mobilio, Marco 357
 Morrelli, Massimiliano 350
 Moustakas, Konstantinos 344
 Muuraiskangas, Salla 27
 Mylonas, Georgios 75, 186

 Napoletano, Paolo 357
 Nesi, Ilaria 75
 Nousias, Stavros 344

 Oğul, Burçin Buket 330

 Paganelli, Federica 75
 Pansini, Giacomo 350
 Pantziou, Grammati 108
 Papeorgopoulou, Penny 283
 Pes, Barbara 318
 Petrak, Björn 324
 Pocero, Lidia 186
 Polito, Massimiliano 350
 Prost, Sebastian 41
 Psaltis, Antonios 283

 Räsänen, Pauli 27
 Reforgiato Recupero, Diego 311, 318
 Reisinger, Michaela R. 41
 Ribeiro, Anabela 296
 Riboni, Daniele 318
 Rizopoulos, Charalampos 283
 Ronkainen, Jussi 27
 Rossi, Claudio 337

 Şahin, Pinar Duygulu 330
 Sarthak 252
 Schrammel, Johann 41
 Seetsen, Ward 139
 Seiderer, Andreas 266
 Shashaj, Ariona 350
 Shukla, Shikhar 252
 Similä, Heidi 27
 Simões, João 296
 Simou, Ioulia 219
 Spirakis, Paul G. 235
 Stuckenschmidt, Heiner 123

 Theona, Iouliani 283
 Trakadas, Panagiotis 235
 Trollmann, Frank 203
 Tsampas, Stelios 186
 Tselios, Christos 344
 Tsybul'nyk, Vladyslav 14

 Valtorta, Jacopo Maria 58
 Van Dantzig, Saskia 1
 Vehmas, Kaisa 27
 Vildjiounaite, Elena 27

 Wegener, Rebekah 370

 Zahariadis, Theodore 235
 Zaroliagis, Christos 108