University of Applied Sciences and Arts
Northwestern Switzerland

Energy saving in smart homes based on consumer behaviour data

Author:
Michael Zehnder

January 31, 2015

Master of Science in Business Information Systems

Supervisors:
Prof. Dr. H. Wache
Prof. Dr. F. Witschel

ABSTRACT

This paper discusses how energy can be saved in smart homes without lowering the comfort of the inhabitants, based on consumer behaviour data only. A recommender system was designed, that suggests actions for inhabitants without the necessity for installing additional devices, executing manual configuration or having any other interaction with the system.

As a consequence of the devastating earthquake and the resulting nuclear disaster that struck Fukushima in March 2011, concerned members of the public and the government agreed on a major reconsideration of the energy policy. However, such a radical rethinking can only be achieved if private households increase their efforts to save energy. Nevertheless, most research approaches conducted in smart homes in the past years, dealt with convenience rather than with sustainability. The aim of this master thesis is to find a way to save energy without causing significant inconveniences for the consumer. Therefore, the following hypothesis was formulated: “It is possible to design a recommender system that can suggest actions in smart homes based on consumer behaviour, which will lower energy usage but not decrease comfort levels”.

The approach followed in this paper, is to mine frequent (and/or periodic) patterns in the event data of the inhabitants electricity usages, recorded by a smart home automation system. These patterns are converted into association rules, prioritized and compared with the current behaviour of the inhabitants. If the system detects opportunities to save energy without decreasing the comfort level, it will send a recommendation to the residents.

Because the most appropriate research design to prove this hypothesis is design science research, the project follows the methodology to design and implement a functional prototype of a recommender system. At the end of the project, the prototype is evaluated in smart homes under real conditions.

The main findings of the project and the concluding field-test of the prototype were:

- The project succeeded in identifying possible actions, which can be recommended in smart homes to lower energy usage in smart homes.
- Investigations showed how patterns in the behaviour data of the inhabitants can be used to trigger these actions at the right moment, to not lower comfort levels for the inhabitants.
- A design has evolved for a recommender system that uses association rules and deterministic finite state machines.
- It was identified, that the confidence and the length of a pattern are significant measures to predict if a suggestion does lower comfort or not.

Overall, it can be said that this master thesis could verify part of its statement: The prototype demonstrated that it is possible to suggest actions that lower energy usage, but do not decrease comfort levels, while using consumer behaviour data as single source. However, besides the useful recommendations, the system did still recommend actions that did not just lower energy usage, but also the comfort level of the inhabitants. The ratio of useful recommendations, which reached little over 11% during the final test of the prototype, must be increased before broader adaption of the system is possible. Nevertheless, the proof of concept provided by the prototype is the first important step for further research in this field.

ADKNOWLEDGEMENTS

The author of this paper would like to acknowledge and thank the following persons for their valuable insights and contributions to this project. Without them, it would not have been possible to bring this project to a successful conclusion.

Supervisors from FHNW:

- Prof. Dr. Holger Wache
- Prof. Dr. Hans Friedrich Witschel

Contacts from Aizo:

- Miguel Rodriguez from Aizo
- Danilo Zanatta from Aizo
- Andrea Helfenstein from Aizo

Everybody who contributed to the evaluation of the system:

- Carla Janzen
- Christian Hitz
- Christoph Hofmann
- Erik Van Dort
- Irmy Hofmann
- Jerome Zuercher
- Paul Buchmeier
- Samuele Di Lernia
- Volker Deckers

Contents

CHAPTER 1 - INTRODUCTION

1.1 INTRODUCTION OF CHAPTER 1

A smart home is defined as a “dwelling incorporating a communications network that connects the key electrical appliances and services” which can be “remotely controlled, monitored or accessed” (King 2003). This definition is correct for most smart homes, containing interacting and connected devices for automation and control purposes, but it does not establish a direct connection to the term “smart”. A home which does act in a smart way, is a system which is autonomously operating, based on artificial intelligence (Blumendorf 2013). However, this definition creates a problem resulting from the difficulty to measure the smartness of such a home. There exists no standardized or common way to evaluate how smart a system is. The recommender system designed in this research project is such an autonomously operating system, but its smartness is evaluated by the energy savings created without loss in comfort.

The purpose of this master thesis is to find a way, to save energy without accepting significant inconveniences to the consumer. Therefore, it uses the data from the key electrical appliances recorded by an existing smart home automation system, to analyse the behaviour in the household. Based on this data, the recommender system will give recommendations how to save energy in the smart home. The recommendations are based exclusively on the existing event data, no motion detection devices have to be installed. Furthermore, no manual settings must be configured by the user; the system will act completely independent. Besides the recommendations, how to save energy, no interaction between the consumer and the system occurs. Furthermore, the recommender should only decide to make an energy saving suggestion to the inhabitant, if no lack in comfort is expected.

The following chapters are going to elaborate, how such a recommender system can be designed and will evaluate, based on a thesis statement, if the recommender will perform the expected results.

1.2 BACKGROUND INFORMATION

The “Energy Strategy 2050” aims to turn over the energy production from big nuclear power plants to many but small renewable energy sources like solar collectors (Swiss Federal Office of Energy 2014). This ambitious goal can only be achieved if private households reinforce their efforts to save energy, which could be achieved by the means of smart home technology.

Home automation technology is becoming as synonymous with sustainability as it is with convenience. However, most research approaches published in the past years dealt more with convenience than with sustainability (Brush et al. 2011). However, there exist several publications about approaches to make smart homes more sustainable, either by improving the sustainability of the technology itself or by using smart home technology to make the home more sustainable (Blumendorf 2013). In this document, the second approach will be discussed. As the title might suggest, a recommender is designed, to make suggestions to improve the smart home’s sustainability but do not lower its convenience.

1.3 PROBLEM IDENTIFICATION

Saving energy in sustainable homes is often linked to high costs, which might not be amortized. Another problem for inhabitants is the amount of time they have to spend for initial configuration and further maintenance of an automation system, which saves energy. A study of long-term use of home automation conducted by Microsoft Research at University of Washington in 2011 revealed the top four barriers that need to be solved, before home automation becomes amenable for broader adoption. The barriers were high cost of ownership, inflexibility, poor manageability and difficulties achieving security. (Brush et al. 2011). While the difficulty of achieving security is not addressed in this master thesis, the project elaborates possible solutions for the three other barriers. Because not only monetary aspects succeed in high costs of ownership but also the time spent for managing the system, the interaction between the consumer and the system must be limited to a minimum. Furthermore, the costs of acquisition are kept low, because the system operates on low-level sensor data, which already exist in most smart homes. Therefore, no further hardware has to be acquired.

Nevertheless, a design for such a system, which gives effective suggestions to save energy, is still to discover. Augusto & Nugent mentioned in their publication "Designing smart homes: the role of artificial intelligence" (2006) that "the relation between the actions performed [...] and the resulting states is a fundamental part of an automated diagnosis system, however [...] is not clear, what system will be most effective in predicting possible outcomes given a trend of behaviour." This problem "poses a real challenge for a diagnostic system to give useful advice whilst minimizing the chances of raising false alarms".

Most smart home systems dealing with artificial intelligence (AI), use motion sensors and person tagging for location prediction of the inhabitants, to solve the problems mentioned. The system designed in this master thesis, deals with the challenge of operating without such enhanced sensors for location prediction. Furthermore, the recommender has no information about the floor plan of the house and the location of the devices. For example if the cooker in the kitchen is on, but the person just has left the room, the recommender will not receive this information. Even if all lights in the kitchen were turned off, the sensor data provided to the recommender do not reveal that the cooker and the lights are in the same room.

However, the assumed independency between the data from the sensors and the metadata about the sensors (e.g. location), will increase the applicability and make the recommender more flexible to use in different locations. The system should adapt over time to the behaviour of the inhabitant and therefore be applicable for a general smart home, where the recommender system can access low-level sensor data.

Another common problem for smart home automation system are multi inhabitant households, because "most of these techniques have been applied in the context of a single resident" (Crandall & Cook 2008). The system developed in this project should also work in smart homes with more than one inhabitant.

The approach followed in this paper, is to mine frequent usage patterns detected in the behaviour data of the consumer and compare it with what is happening in the house currently. In the kitchen example provided above, this would mean, that the recommender should still be able to detect the unattended cooker and suggest turning it off, because this situation represents a chance to save energy. The recommender should detect in the historical data provided, that when the cooker is on, usually also the light sources in the kitchen are activated. Due to the fact, that the current state is very abnormal, the recommender could detect this anomaly and suggest the inhabitant to turn off the cooker.

1.4 THESIS STATEMENT

The problem identified in the previous chapter led to the following thesis statement:

It is possible to design a recommender system that can suggest actions in smart homes based on consumer behaviour, which will lower energy usage but not decrease comfort levels.

1.4.1 Definition of terms

A r**ecommender system** is a software tool or a technique providing suggestions which are useful to its user (Ricci, Rokach & Shapira 2011). In this thesis statement, the term refers to a software system that is providing recommendations related to a decision-making process, such as how to save energy for its consumer.

A **suggested action** is a targeted activity contributed by the system, which can be (manually) performed by the consumer.

The reasoning is **based on consumer behaviour** only, no further interaction between the consumer and the system can occur. The exclusive source of input data for the recommender system is event data, caused by the consumer and recorded by smart home sensors.

To **lower energy usage,** stands for reducing the power consumption in the smart home. The savings are not measured in a monetary way, but only by the fact that a lower amount of energy is consumed by the smart home.

No **decrease in comfort levels** denotes that the energy savings do not go along with negative side effects for the consumer. The savings must not result in any change of behaviour for the inhabitant.

1.5 RESEARCH QUESTION AND OBJECTIVES

To prove or reject the thesis statement, research questions are answered. The research questions were derived from the statement and defined for this study as follows:

RQ1: What are possible actions, which can be suggested in smart homes to lower energy usage?

RQ2: Which types of behaviour patterns are needed to suggest actions?

RQ3: What algorithms can be used to find relevant patterns in the behaviour data?

RQ4: How can such a recommender system be designed?

RQ5: How can be evaluated, if a suggested action does not decrease comfort levels?

RQ6: Do the suggested actions decrease comfort levels?

1.6 RESEARCH OBJECTIVES

In order to answer the research questions, the following research objectives have been formulated:

1. To identify possible actions, which can be suggested in smart homes to lower energy usage.
2. To define which types of behaviour patterns are needed to suggest actions.
3. To find algorithms which can be used to find relevant patterns in the behaviour data.
4. To design a recommender system.
5. To evaluate if the suggested actions do decrease comfort levels.

1.7 DELINEATION AND LIMITATIONS

The recommender system designed in this master thesis is different from other smart home automation systems, which can be found in literature, in terms of goals, reasoning, technology and evaluation. The recommender is designed that it deliberately operates without the need for enhanced sensors, providing visual signals. Neither motion and occupancy sensors nor tagged gloves and RFID-tags are used to determine the location of the inhabitants, since the installation of such devices is expensive and could affect the wellbeing of the inhabitants. Furthermore, the reasoning to provide the energy saving suggestions will not depend on user interaction, but only of the data provided by the sensors. Moreover, the aims of the system do contrast to similar systems: It is not desired to increase the comfort or the quality of life for the inhabitants, the primary goal is to build an autonomously acting recommender system that can help to save energy. In contrast to most other home automation concepts, the evaluation of the system is based on a functional prototype of the recommender, which is implemented and tested in the field to generate hard evidence.

The recommender system developed in this research project is not the sole solution to resolve this problem; other techniques may lead to similar results.

1.8 STRUCTURE AND CHAPTER OVERIVEW

This master thesis comprises seven chapters. Figure 1 outlines the structure of this document and in which chapters the research questions will be raised and answered.

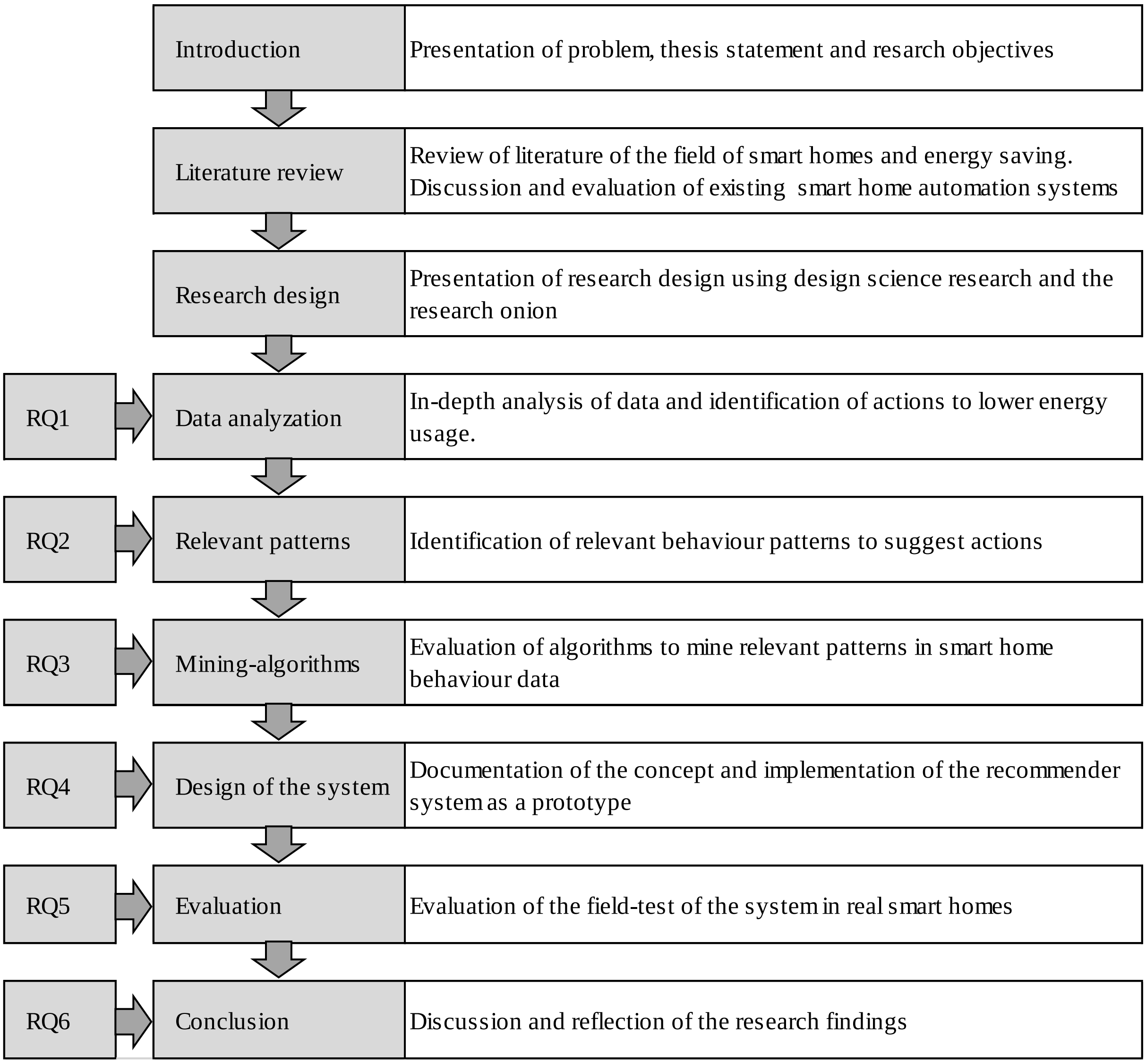


Figure 1 Structure of this document

1.9 SUMMARY OF CHAPTER 1

In this first chapter, the general aspects of this master thesis are addressed. This includes the thesis statement and the research questions. In the last chapter of the introduction, a chapter overview is presented to provide insight in the structure of the document and to give information about the content of the following chapters.

The problem to solve in this master thesis is to find a possibility to save energy by using only the data available, without investing time or money into the manual configuration of the system or the installation of additional sensors. To prove or reject the thesis statement, the system is evaluated in a field-test using real data.

CHAPTER 2 - LITERATURE REVIEW

2.1 INTRODUCTION OF CHAPTER 2

The second chapter is looking at literature, regarding different topics covered by this research project. Concepts and methods from the field of smart home automation systems are discussed and existing approaches are analysed. The goal is to extract proven ideas of such systems and combine, enhance and reuse them in a way that will allow to make recommendations for energy saving. To assess the existing smart home automation systems found in literature, criteria were defined. These criteria resulted (in most cases) directly from the thesis statement and describe the requirements for the recommender system defined by the statement. Because the thesis statement does not set specific limits for all aspects, further criteria were exposed in interviews with experts from the field of smart home automation systems. The criteria for the recommender system were defined as follows:

- Because an evaluation under real conditions is required to prove that no losses in comfort occur for the consumer, the system must be optimized for the use in smart homes. This includes:
    - Being able to deal with a continuous sequence of sensor data
    - Being able to deal with a large amount of sensor data
- To meet the requirement of reducing the power consumption, the system must provide energy saving features.
- To recognize the behaviour of the inhabitants, the system provides detection for the following patterns:
    - Frequent patterns
    - Periodic patterns
- Since recommendations are time critical and only useful in a certain period, the system must provide real time behaviour reasoning.
- Because the behaviour of the consumer(s) can change over time, the system must be able to adopt to changing inhabitant patterns.
- The system does not depend on labelled input data (unsupervised learning).
- The system can operate without any form of presence sensors (motion, occupancy, person tagging, etc.) because this would make additional installations necessary and therefore lower the comfort of the inhabitants.
- To compete successfully with the conditions provided, it must provide multiple resident capability. The system is evaluated in average Swiss households, where the average number of inhabitants per household is above two. One-person households represent less than 35% of all private households (Federal Administration 2012).
- The system can operate without feedback of the inhabitants, because any interaction required would lower the comfort of the inhabitants.

To evaluate the criteria mentioned, the following subchapters will address the main concepts of smart home automation systems. In the final subchapter, the findings are summarized.

2.2 SMART HOMES IN GENERAL

Because smart homes have become a very active and firmly established research topic, many publications about the subject can be found. The area is fast developing and attracting synergies of several areas of science, for improving human’s quality of living. To address the needs within a smart home environment effectively, multidisciplinary collaboration among different research topics is necessary.

Richard Harper, who examined the field of smart technologies for private homes, wrote that the way a house was built or the environment friendly features it contains will not make it a “smart home”. But “what makes it smart, is the interactive technologies that it contains” which could help to realize “the dream of a home, which can proactively help its inhabitants” (Harper 2003). Although the term “smart home” was officially mentioned the first time in 1984 by the American Association of House Builders, the first “wired homes” were actually built years before and were essential, to develop what is meant by smart homes today. The expectations at smart homes back then were to increase productivity in domestic work, with little involvement of the user into the technology. (Harper 2003)

For a long time, research conducted about smart homes has been largely focused on hardware solutions. Nowadays, the term mainly addresses the integration of information technology into domestic buildings. Security, health care, energy efficiency and increasing the comfort of the inhabitants are the main research topics in the field of smart homes. Interconnections between devices and advanced control over lights, entertainment and multimedia devices for increasing comfort are proposed in several publications (Jakkula et al. 2006; Björkskog 2007; Nowak & Urbaniak 2011). Furthermore, in the area of security, remote information and intervention systems are developed, to increase the control inside the house while the inhabitant is not present. Another application of security systems mentioned in literature is presence simulation (Clinckx 2013). Because a presence simulator will not depend on complex rules to pretend attendance based on historical data, there is not much space for further research in this area and similarities between such systems and the recommender system developed in this research project are difficult to find. However, in other areas, there are still opportunities for more research to be conducted. In support for dependent people and home care, there are two different problems to be solved. At first, such a support system should notice when an emergency occurs inside a house and its second task is to increase the comfort for the person in need of care. However, an increase of convenience is not only interesting for elderly people; it is also a desire of average inhabitants.

2.3 SAVING ENERGY IN SMART HOMES

Already in 2003, Harper acknowledged, that “smart house technologies that most people are pleased with, are connected with saving energy or money” (Harper 2003). But especially since the Swiss Federal Institute pictured the environmental vision, to reduce the continuous overall average energy usage per person to not more than 2000 watts by the year of 2050, without lowering the standard of living, the idea of saving energy became more than just a monetary aspect (Energie Schweiz für Gemeinden et al. 2012).

In literature, many approaches for energy saving in buildings using smart technology can be found. Recently a publication “Building sustainable smart homes” (Blumendorf 2013) summarized, that the increasing usage of information technology in our homes and the increasing demand for sustainable life style are two contradictory concepts and raises the question, of how to combine these two trends. One of the proposed challenges, to “make users aware of the hidden details of their current behaviour” is addressed in the research project at hand.

Most approaches for energy saving in smart buildings found in literature, are focused to lower the energy consumption of heating, ventilation and air conditioning (HVAC) devices, such as the domestic heating system (Villar et al. 2009), air-conditioning (He 2010) or both of them (Nowak & Urbaniak 2011; Inji et al. 2011). Others do not directly address to lower the consumption of such devices, but provide enhanced monitoring and controlling (Jahn et al. 2010). Most of these projects use a broad variety of sensors to measure humidity and temperature and process the data by a fuzzy controller (Villar et al. 2009; Nowak & Urbaniak 2011) to distribute the energy. The project at hand aims to disclaim enhanced temperature and humidity sensors and consequently there are insufficient similarities between these projects and the current problem. Others who were also anxious to “minimize the domestic energy waste” and identified two more areas to incorporate energy management features: “lighting and home appliances” (Inji et al. 2011). In simulations over different scenarios in synthetic data, they discovered that potential for energy savings in private households is nearly 30 percent. To trigger lighting devices, they examined occupancy sensors, which would capture the individual by infrared heat detection. Because the thesis at hand does not make use of such sensors too, it will be difficult to reproduce savings of this magnitude.

Another interesting project, which occurs in several publications, is the MavHome (Managing an Adaptive and Versatile Home) project, a multi-disciplinary research project at Washington State University and the University of Texas. The MavHome is an extensive project, where many researchers from different topics contributed. Its goal, to maximize the comfort, while minimizing the operational cost is very similar to the one at hand. Many publications contributing to this project can be found on the corresponding website (MavHome 2003). Since some of the project examine systems to predict the actions of the inhabitants of the house, they will be scanned more closely in the impending chapters.

2.4 SUPPERVISED AND UNSUPPERVISED LEARNING

Learning about the behaviour of an inhabitant is essential for a recommender in a smart home, in order to self-tune the system and to support its independence and personalization. The system should be acquainted with the habits of the inhabitant and continuously learn and improve them. With the intention to achieve that, the system must find reoccurring patterns in the smart home event data and learn these frequent “activities”.

In machine learning, there can distinguished between two fundamentally different learning paradigms: supervised and unsupervised or self-organized learning (Augusto & Nugent 2006). In supervised learning, the algorithm is given samples of the different types of classes, which can be found in the data. The algorithm uses this information to segment the unseen data into the appropriate classes. In unsupervised learning, no labelled examples are provided. Many supervised activity recognition methods for smart homes have been published in the literature in the last years (Rashidi & Cook 2013). Supervised learning is used in methods like decision trees (Augusto & Nugent 2006) Markov models (Gopalratnam & Cook 2007) and dynamic Bayes networks (Du et al. 2006). Bayes classifiers can detect the activities that correspond to the sensor values, observed with the highest probability. Although these classifiers rely on conditional independence of the events, the classifiers achieve good accuracy with large amounts of training data (Rashidi et al. 2011).

However, these methods have the major disadvantage that labelled data is required to train the algorithm. The manual annotation of human behaviour data, based on low-level event sensor data is a very tedious and time-consuming task and can additionally limit the scalability of the system. Moreover, the authors of “A Method for Mining and Monitoring Human Activity Patterns in Home-based Health Monitoring Systems” examined, that a postulation for consistent predefined activities does not work in reality, because not every inhabitant will perform the same. Therefore, labelled training data can only be used for the household they were made for. (Rashidi & Cook 2010).

The following diagram Figure 2 illustrates the differences between supervised and unsupervised methods for smart home automation systems.

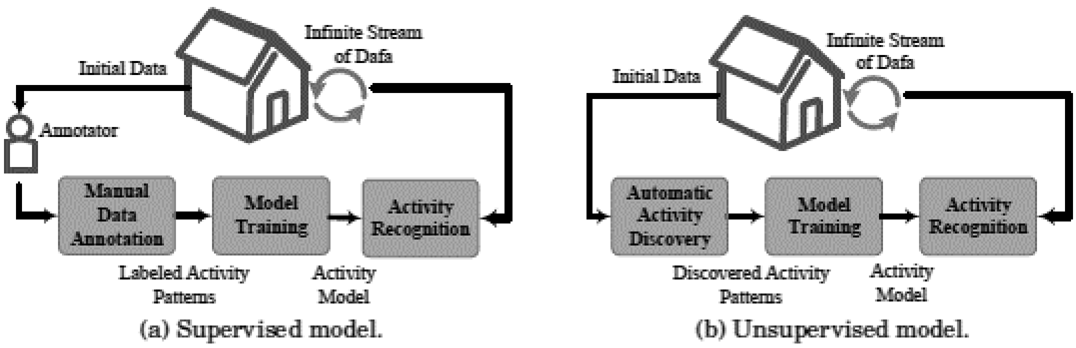


Figure 2 Activity recognition models (by Rashidi & Cook 2010)

Since the goal of this research project is to save energy without negatively influencing the comfort level of the inhabitants, the unsupervised approach seems to be more suitable. In contrast to supervised methods, it requires no training, examples or labelled data. The user is not obliged to scan his data for activities because this would decrease the comfort, the system should find the patterns autonomously.

2.5 MINING SEQUENTIAL DATA IN SMART HOMES

Sequential pattern mining has been studied for almost two decades (Manilla et al. 1995; Srikant & Agrawal 1996) and various approaches have been published for finding sequential patterns in data (Manilla et al. 1995; Srikant & Agrawal 1996; Chen et al. 2012). These methods presuppose that the input is given as transactional data. However, in a smart home environment, this is usually not the case. Sensor data has no boundaries, which separate the different events into activities; they provide just a stream of data. Techniques such as the Apriori algorithm (Srikant & Agrawal 1996) were successfully proposed to group such event data together into activities. In the publication of Achar, Laxman and Sastray, a unified view of Apriori-based algorithms is drawn (Achar et al. 2011).

Finding frequent and periodic patterns in sensor event data, is a data-mining subfield usually referred to as “discovering frequent episodes in sequences” (Manilla et al. 1995; Mannila et al. 1997), “sequence mining” (Youngblood & Heierman 2005; Jain et al. 2006; Rashidi et al. 2009; Achar et al. 2011; Rashidi & Cook 2013) or “activity recognition” (Du et al. 2006; Crandall & Cook 2008; Nazerfard et al. 2011; Chen et al. 2012; Krishnan et al. 2013; Rohini & RajKumar 2014). The starting point for this research was the pioneer work of the Apriori algorithm, which was later implemented and enhanced by other projects. In most of these publications, the differentiation between frequent and periodic patterns is made, because both are desired for later automation.

2.5.1 Frequent patterns

The core functionality of a pattern-mining algorithm is to find frequent pattern. In literature, many approaches for frequent pattern mining in sequential smart home data can be found. Most of them are similar to the Apriori algorithm and take a bottom-up approach using a sliding window (Youngblood et al. 2005; Rashidi et al. 2009; Rohini & RajKumar 2014).

By pruning the infrequent patterns, only the most relevant variations of a pattern will be returned. Through reducing the number of irrelevant variations, also the need to manually configure properties for the mining is lowered (Rashidi & Cook 2013). The percentage of the most frequent sensor events, the support threshold used to mine frequent events or the number of activities that should be returned, are examples for such properties.

2.5.2 Periodic patterns

In smart home automation systems, it is not only important to find frequent activities, but also those patterns that occur the most regularly in predictable periods. If such a system would ignore periodicity and only rely on frequency to discover patterns, many periodic events might be discarded.

The Apriori algorithm in its original form is not able to mine periodic patterns in sequential data and was therefore improved by different people for the use in smart home automation systems, for example for the CASAS System (Rashidi et al. 2009). In this approach, the algorithm starts by mining patterns with a minimum length of two activities, which is extended until the algorithm is not able to find frequent patterns anymore (Figure 3). For every pattern, a tentative periodicity is calculated and then compared to the other instances found in the event data, to check if the average periodicity does apply or not. If not, the pattern is not considered as an instance of the same pattern and stored separately. A threshold (grace period) can be defined, that distinguishes if a periodicity is still within tolerance or not. All patterns with one or more periodicities can be considered as periodic.

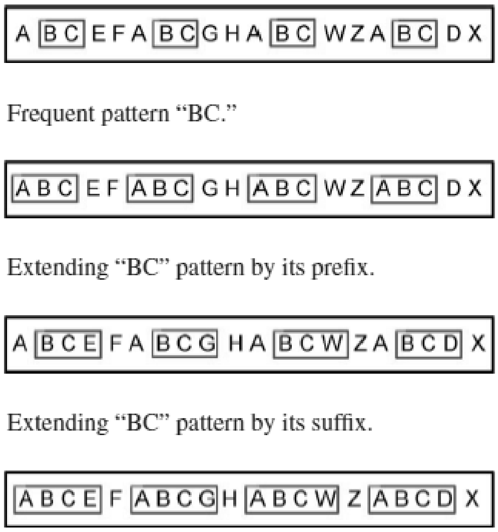


Figure 3 Finding periodic patterns (example of the CASAS System by Rashidi et al. 2009)

Not in all publications, the detailed procedure, how periodic patterns are mined is explained. However this approach can be considered as a model, how most systems mine periodic patterns (Li et al. 2010; Chen et al. 2012; Rohini & RajKumar 2014).

2.6 CLUSTERING SEQUENCES

An approach often mentioned in literature is to identify pattern clusters that represent a set of activities. The resulting clusters represent activities that can be stored and tracked in real time data. Clustering is executed after the sequence mining and results in a compressed representation of the activities. Even though the patterns were already grouped by the mining algorithm in sense of structural similarities, the clustering can improve the grouping by also concerning start time, duration and other similarities. Clustering also addresses the problem of discovering too many similar patterns, when the large number of patterns make it difficult to analyse the underlying ideas. (Rashidi & Cook 2013)

Standard k-means clustering is one method to group the set of discovered patterns into clusters. It is a clustering method using vector quantization, adopted from signal processing, that is popular for cluster analysis in data mining and therefore often used in smart home automation systems (Augusto & Nugent 2006; Ricci, Rokach & Shapira 2011; Rohini & RajKumar 2014). It aims to partition n observations into k clusters by defining a method for determining cluster centroids. In literature, several methods have been reported, such as the ROCK algorithm by Noh et al. (2009) and the CLUSEQ algorithm by Yang et al. (2009), which do only consider symbolic sequences, with no features attached to them. In contrast, sequential smart home data does not only consist of strings, it has also additional information, such as time, associated.

A refined clustering algorithm for smart home automation systems was published by Rashidi & Cook (2011). It combines several activities together to form a new activity, which is called “extended state”. For example, if a sensor in the living room is triggered several times in a row without another sensor event interrupting the sequence, the events are combined into one event with a longer duration. Due to this clustering functionality, a more compact representation of activities is created and as a result, it is simpler to compare similar activities. Two other methods that are used to compare similarities of sequences are edit distance (Manilla et al. 1995; Rohini & RajKumar 2014) and longest common subsequence (LCS) (Sequeira & Zaki 2002) for simple sequences. However, these methods are not sufficient to address the clustering problem in sequential data, because they do not use the temporal information, which would give information about the order of the events.

2.7 SELF-ADAPTING SYSTEMS

Rashidi wrote in one of her publications about smart home systems, that “considering the fact that humans usually change their habits and activities over time, adaptation is a crucial part of a smart-environment solution.” (Rashidi et al. 2009). She proposed a system for automation of sequential complex activities, which adapts to the user’s preferences by using contextual information, such as temporal data and start-up triggers. In contrast to most other systems in literature, where the inhabitant must guide the system by providing explicit feedback, their automation system (CASAS) can automatically discover and adapt to changes in the patterns of the resident.

In the algorithm of the CASAS system, for every successfully mined pattern, an attribute is maintained, which reflects the degree to which it can be considered for automation. The value can be increased or decreased through a compensation or a decay effect. If the value drops below a certain threshold, the pattern is discarded from the model. “The decay effect allows for those activity patterns that have not been perceived over a long period to descend toward a vanishing value over time or, in an intuitive sense, to be forgotten.” (Rashidi et al. 2009). At that time, this feature was pioneering and it was adapted for other smart home automation systems over time (Nazerfard et al. 2011; Alam et al. 2012).

2.8 MULTIPLE RESIDENTS

For most approaches of smart home automation systems, the problem of multiple residents has been completely ignored; they have been applied in the context of a single resident. However in few publications, the problem of multiple residents is identified (Rashidi et al. 2009; Krishnan et al. 2013) but not considered for problem solving, but referred to others, who proposed a solution for this problem (Crandall & Cook 2008). Crandall and Cook found a way of attributing events to individuals in multi-inhabitant environments. Unfortunately, the solution proposed is using a supervised approach (naïve Bayesian classifier), which requires labelled training data to identify the resident responsible for a unique sensor event. Therefore, this approach cannot be adopted in the research project at hand.

2.9 ACTION PREDICTION FOR SEQUENTIAL DATA IN SMART HOMES

Another approach used to prefigure the behaviour of smart home inhabitants, found in literature is action prediction. Instead of using reoccurring patterns to assume upcoming activities, action prediction uses the possibilities for the next occurring events for event forecasts. Prediction is a heavily researched topic in artificial intelligence (AI). To predict events in smart homes, other areas where events in sequential data were anticipated, serve as example. For instance ONSI, a system to predict Unix command lines (Korvemaker & Greiner 2000) and IDHYS, an environment which offers to complete repetitive tasks (Ruvini & Dony 2000), were mentioned to pioneer in prediction for sequential data.

Two Markov based approaches that can be found in literature are the Active LeZi (ALZ) algorithm (Gopalratnam & Cook 2007) and the SHIP (Smart home inhabitant prediction) algorithm (Das et al. 2002). Because the interaction between the inhabitant and the devices is regarded as a Markov chain of events, these algorithms employ this stochastic model to predict the next action optimally. The algorithm identifies all sequences ending with the most recent event and weight them according to number of occurrences and length of the sequence. Both algorithms were tested either on data collected in a real smart home (ALZ) or synthetic data (SHIP) and scored 47% respectively 53,4% accuracy for predicting the next event. However, the limitation of this algorithm is, that it does not consider the time, when such an event happened. Since the algorithm can only predict the next event in a sequence and not the time it occurs, the application in a smart home is only possible within a limited basis.

2.10 SUMMARY OF CHAPTER 2

In this chapter the literature regarding smart home automation systems, their techniques and their potential to suggest recommendations for energy saving was discussed. To achieve energy saving without a lack in comfort for the inhabitant, the system must be able to predict the inhabitants’ behaviour. Therefore, the smart home system has to recognize the usage patterns of the inhabitants depending various appliances. It was recognized, that manually specified preferences, usage patterns is no acceptable approach, and that therefore unsupervised machine learning is needed. Furthermore, working concepts for frequent and periodic pattern mining were found, which can be adopted. Also an existing idea to make such a system self-adapting was explained and the challenge of systems dealing with multiple residents analysed.

Different systems and algorithms where examined and evaluated according to the criteria defined in the introduction (chap. 2.1) of the literature review chapter. The results of the evaluation are presented in a summarized way in the following table, Table 1.

	ALZ	BAS	CASAS	COM	DVSM	IntelliDomo	PUBS	SPEED	TEREDA	WSDD
Optimized for sequential smart home data	*[1]	+	+	+	+	+	+	+	+	+
Tested (successfully) with real smart home data	-	-	+	+	-	+	+	+	+	+
Provides energy saving features	-	+	-	-	-	-	-	-	-	-
Frequent behaviour pattern detection	+	-	+	+	+	+	+	+	-	+
Periodic behaviour pattern detection	-	-	+	+	+	-	-	+	-	+
Real time reasoning	+	+	-	-	+	-	+	-	-	-
Adapting over time	+	-	+	+	+	-	-	+	+	-
Unsupervised approach	-	-	+	+	-	-	+	+	-	+
Operating without motion sensors	+	+	-	-	-	+	-	+	+	+
Multiple resident capability	?	-	-	?	?	?	?	?	+	?
No user interaction necessary	-	-	+	+	-	-	-	+	-	+

Table 1 Evaluation of existing smart home automation systems found in literature

+ = criteria fulfilled, * = criteria partially fulfilled, - = criteria not fulfilled, ? = not assessable / unknown

Abbreviation	Name	Author	Additional Information
ALZ	Active LeZi	Gopalratnam & Cook (2007)	[1] the original ALZ algorithm was not designed for smart homes, but adapted by the MavHome project for this purpose.
BAS	Building Automation Systems	LeMay et al. (2009)	
CASAS	Continuous Adaptive Smart home Access System	Rashidi et al. (2009)	
COM	Continuous varied-Order Multi Threshold activity discovery	Rashidi & Cook (2013)	
DVSM	Discontinuous Varied-Order Sequential Miner	Rohini & RajKumar (2014)	
IntelliDomo	Intelligent System using a domotic database	Boton-Fern´ndez & Lozano-Tello (2011)	
PUBS	Patterns of user behaviour system	Aztiria et al. (2011)	
SPEED	Sequence prediction via enhanced episode discovery	Alam et al. (2012)	
TEREDA	Temporal Relation Discovery of Daily Activities	Nazerfard et al. (2011)	
WSDD	Window Sliding with De-Duplication	Schweizer (2014)	

Table 2 Abbreviations for the systems evaluated in Table 1

Even though the literature review showed that a lot of research was and is still being done in the field of smart homes, energy saving and automation systems, there are unsolved problems remaining. The most important aspects, which must be solved, are:

- A recommender system, that provides concrete suggestions to the inhabitant, how energy can be saved. No other system was found in literature that could serve this purpose.
- A system that uses data, which is exclusively collected from the usage of real appliances. Most projects reviewed, accessed additional sensor data such as motion, occupancy, temperature and humidity sensors. It remains to be examined, if a home automation system without access to such additional sensors can still provide useful suggestions for energy saving.
- A system that is acting autonomously, where besides the recommendation sent to the user, no direct interaction between the system and the inhabitant occurs.
- A system which is using real time event data from real smart homes and which was also evaluated in (a) real smart home(s). Consequently, such a system can achieve a higher significance regarding real-world applicability.

CHAPTER 3 - RESEARCH DESIGN

3.1 INTRODUCTION OF CHAPTER 3

The purpose of this chapter is to describe the particular research plan and design used for this research project. The choice of the suited research design is important for the success of the research project and has to be considered carefully.

3.2 RESEARCH STRATEGIES

Johnson & Clark (2006) emphasized the importance of the choice of the research strategy, because it has not only a significant impact on what the research project will consist of, but it will also help to understand what exactly is investigated. Saunders et al. (2009) described seven possible research strategies:

- experiment
- survey
- case study
- action research
- grounded theory
- ethnography
- archival research

They mentioned that what is most important is not the name of a particular strategy, but whether it will allow answering the research questions or not. Thus, the choice of the research strategy will be guided mostly by the research objectives. Subsequently the research strategy defines the collection of techniques and analysis procedures, which are applied during the research project. Finally, it is important to mention that these strategies are not exclusive; it is possible to use one as part of another.

In addition to these research strategies, there is design research. Design research is a set of analytical techniques and perspectives and was described by Vaishnavi & Kuechler Jr. (2007). It is often used for performing research in “Information Systems”, because it is the most suited approach to develop a solution for a problem by designing an artefact, which can be evaluated.

3.3 THE RESEARCH DESIGN OF THIS MASTER THESIS

Because the goal of this research project is to design and implement a functional prototype of a recommender system, which is able to make suggestions to save energy in smart homes and to evaluate the prototype in a real smart home, the most appropriate research design for this master thesis is design science research.

Design science research is structured in five phases called process steps. These steps are illustrated in the Design Science Research Model (Figure 4). The model is an adaption of a design process model developed by Takeda et al. (1990) and attaches significance to the activities carried out in the different phases.

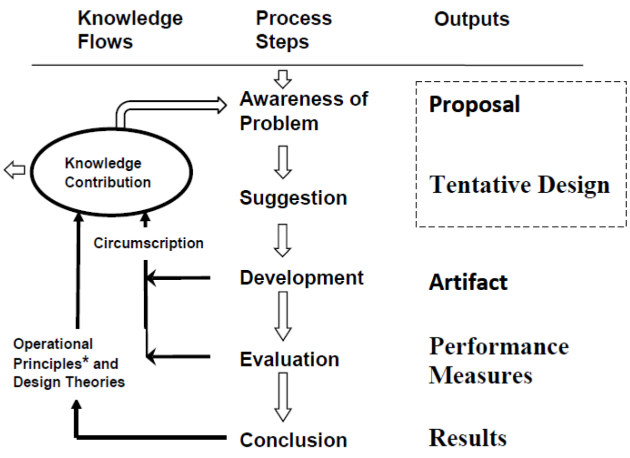


Figure 4 Design Science Research Model according to Vaishnavi & Kuechler (2007)

To process design science research according to Vaishnavi & Kuechler (2007) the steps of the "Design Science Research Model" and the research goals defined for the project should be harmonized. The following table (Table 3) will link the phases to the goals and explain which activities inside a phase will help to answer the corresponding research question.

Phase	Research question	Activity
Awareness	What are possible actions, which can be suggested in smart homes to lower energy usage?	Literature research, interviews with experts, analysis of the underlying data
Suggestion	Which types of behaviour patterns are needed to suggest actions? What algorithms can be used to find relevant patterns in the behaviour data?	In-depth analysis of data, identification of sample patterns for energy saving, analysis of existing algorithms
Development	How can such a recommender system be designed?	Analysis of the runtime environment and real-time sensor data, implementation of a prototype
Evaluation	How can be evaluated, if a suggested action does not decrease comfort levels?	Design setup for a field-test in real smart homes
Conclusion	Do the suggested actions decrease comfort levels?	Consolidation of the results

Table 3 Research strategy according to the phases of the Design Science Research Model

3.3.1 Awareness of Problem

The goal of the awareness phase is to collect as much information as possible about the problem explained in "Problem identification" (chap. 1.3). In order to bring together the information needed to get aware of the problem, different techniques will be applied: review of related literature, interviews with experts and the structural analysis of the historical smart home event data provided. To cover all the important areas of literature, the literature review (chap. 2) will be structured as follows:

- The first part should provide general information about smart homes and summarize the initial but also the most recent research, conducted in this area (chap. 2.2).
- The second part will concentrate on recent publications about energy saving in smart homes (chap. 2.3).
- In the last and most specific part, similar smart home automation systems will be examined. The goal of this part is to create a catalogue of requirements (Table 1) that a smart home automation system should satisfy, to solve the problem identified in this master thesis (chap 2.4 ff.).

By means of interviewing experts in the field of smart home automation systems (digitalSTROM), innovation for problem solving can be accomplished. It is expected that the experts can deliver valuable information to the following subjects:

- Possible actions to lower energy in smart homes (research goal 1).
- Details about technical environment where such a recommender system could run.
- Precise information about sensors (types, measurements, latency, output values, etc.).

The analysis of the historical data collected in real smart homes, provides the last part of information collected for the awareness phase. The concrete smart home event data is provided by Aizo, a Swiss company that develops and sells the smart home automation system "digitalSTROM". The data is available as secondary data for this project. It was chosen, because it satisfies all of the following criteria:

- The data offers hard evidence collected in real-life smart homes.
- There is enough data for a detailed analysis (several years of data for 34 smart homes).
- The data is very recent and can be updated on request.

The output of this phase is a formal proposal (chap. 1-3 of this master thesis).

3.3.2 Suggestion

The suggestion phase follows immediately after the proposal and should develop an idea for the solution of the problem identified in the proposal. This is a creative step, where new functionality is envisioned, based on existing elements.

The first part of this phase will consist of a detailed and accurate data analysis. The quantitative data is provided as an export of a relational database. It consists of about one gigabyte of numerical data. The data analysis is a fundamental step to suggest a tentative design, because it will not only provide information about what can be achieved with the data, it will also set the limits and provide clarity what is not possible. The data analysis will contain the following steps:

- Understand the structural relations between the tables.
- Initial analysis and visualization of the data.
- In-depth analysis for event data.
- Find samples for patterns where energy could be saved.

Based on the data analysis and the sample patterns identified, the second step is to define the types of behaviour patterns, which could be used to suggest actions to save energy (research goal 2). Since the data provided will contain several years of data of more than 30 smart homes, it is pointless to define patterns, which do not occure within the historical data. The sample patterns should match as many as possible of the following criteria:

- They occur frequently (significantly more often than other patterns or single events).
- They occur in periodic manner (reoccur in frequent periods).
- They represent a considerable potential for energy saving (defined by research goal 1).

The last step of the suggestion phase is to suggest a design of a recommender system (a completely new approach or a modified version of an existing automation system) that can be used to mine relevant patterns (defined by research goal 2) in the behaviour data, to suggest actions for energy saving (research goal 3). The selection will be based on the following facts:

- Evaluation criteria table for the suitability of existing automation systems (Table 1)
- Suitability to detect relevant patterns (with energy saving potential) defined by research goal 2

The output of this phase will be a suggestion for a tentative design.

3.3.3 Development

In the development phase, the tentative design is developed further and implemented to a functional prototype, which can be applied to the available data (research goal 4). The implementation itself can be ordinary; the novelty lies in the design and not in the way of construction. It is crucial to develop the prototype according to the specifications in the previous phase, to satisfy the research goals 1-3. In addition, it is important to consider the following points, to prepare the artefact for the evaluation phase:

- The prototype must run in an environment specified by the experts from Aizo in the awareness phase in order to make a field test possible (Table 3).
- For a successful evaluation, the prototype must be able to process the real-time data from the sensors in the smart homes, specified by the experts in the awareness phase (Table 3).

The output of this phase will be an artefact, represented by a functional prototype of the recommender system which can be evaluated in a field-test.

3.3.4 Evaluation

In this phase, the artefact will be evaluated according to thesis statement and the criteria defined in the research questions.

In order to prove or reject the thesis statement, the artefact will be evaluated in a field test. Because of research goal 5, it is not possible to evaluate the prototype just on a basis of historical sample data. To examine, if the comfort of the inhabitant is lowered due to the recommendations made by the system, immediate user feedback to each suggestion is necessary. For this reason, Aizo agreed to provide real smart homes for a field test. The smart homes for the field test are a subset of the households, in which the historical data was collected and will therefore operate with the same sensors. The prototype will be installed in 5-10 smart homes with volunteer inhabitants. It is likely, that the test homes are habited with different quantities of people, in order that the system can be evaluated for single as well as for multi-habitant functionalities.

During the evaluation phase, no interaction between the inhabitants and the system should occur (except the recommendations provided). It is proposed that the suggestions will reach the inhabitants on a mobile device. After a recommendation was made, the inhabitant can evaluate if the suggestion was "useful" and would not negatively influence his comfort level or if the suggestion was inappropriate. This data will be used to answer the research questions and to prove or reject the hypothesis.

In design science research, the initial hypothesis is rarely proven by the first prototype. Often the evaluation results and the information gained during the tests are brought together and fed back to improve the artefact, starting with a new suggestion. (Vaishnavi & Kuechler 2007). Therefore, two evaluation phases are planned. During the first phase, the goal is to make a large quantity of suggestions in order to gain as much experience about the system as possible. These experiences are analysed and considered for the second phase where the goal is to increase the ratio of useful recommendations.

The output of this phase will be concrete performance measurements of the recommender system.

3.3.5 Conclusion

The conclusion is the final phase of the cycle, where the results are consolidated and communicated. Since communication is very important in research, the contribution of knowledge is essential for future work (Vaishnavi & Kuechler 2007).

3.4 SUMMARY OF CHAPTER 3

A summary of the research design of this master thesis, will be drawn by using the research onion from Saunders et al. (2009).

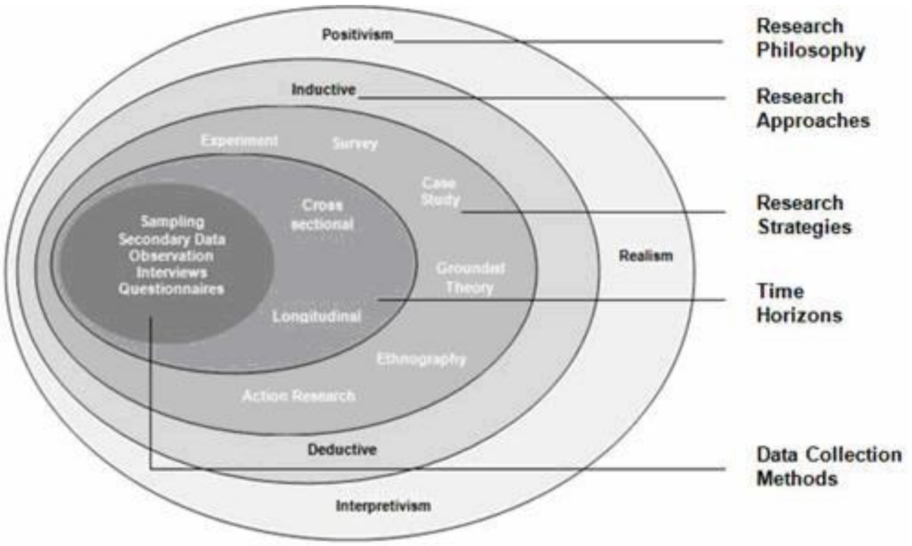


Figure 5 The research onion from Saunders et al (2011)

3.4.1 Research Philosophy

The goal of this master thesis is to gain information that justifies the belief, that it is possible to save energy in smart homes under given circumstances. This belief can only be justified by empirical evidence gained by observation in a field test. Because of this, the research philosophy of this master thesis can clearly be defined as positivism.

3.4.2 Research Approach

The classification of this research project in an inductive or a deductive approach is rather difficult, because both are applied in different stages. In the awareness phase, this research project is following a bottom-up approach: To get an understanding of the problem at hand, the sample data is analysed to derive a theory. The foundation for this problem solving process is existing knowledge and theories from literature and experts. This approach is called abduction. Whereas induction tries to find an argument to explain the data analysed, a solution drawn abductively is only supplying a warrant that enables to express a theory from the data (Douven 2011).

However, this research project is not following the abductive approach exclusively. The derived theory, which is made substantial in form of a recommender system, is benchmarked in a deductive approach in the second part of the project. These two phases are shown as cognitive processes in the following Figure 6.

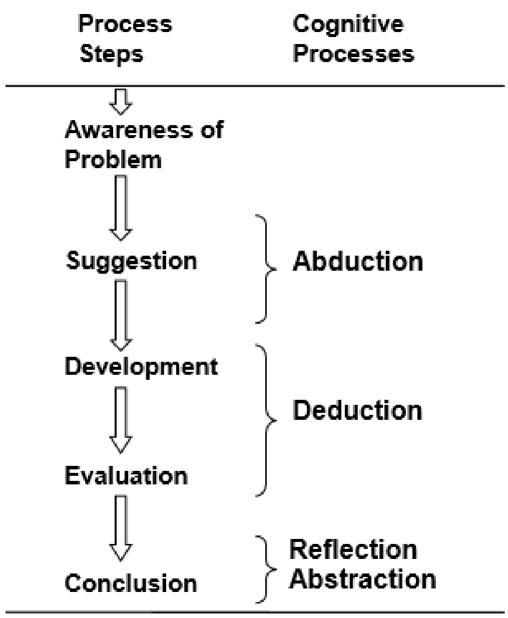


Figure 6 Cognition in the Design Science Research Cycle by Vaishnavi & Kuechler (2007)

3.4.3 Research Strategy

The research strategy of this project can be assigned to design science research.

3.4.4 Time horizon

The sample dataset for the training of the recommender system provided by Aizo contains almost two years of data. For the discovery of frequent and periodic activity patterns, it is anticipated that the more training data is provided, the better results can be expected. However, user behaviour can change over time and there could be historical outliers or noise, which will mislead the mining algorithms. To find the best time spans for the mining process will be an essential challenge during the research process. However it is given, that the spans will comprise rather months then days. This will allow making a meaningful statement regarding the possibility of finding the activity patterns in an autonomous way.

CHAPTER 4 - DATA ANALYSIS

4.1 INTRODUCTION OF CHAPTER 4

The following chapter takes place in the awareness phase of the project and addresses the need for an in-depth analysis of the smart home behaviour data provided by Aizo. An analysis of the data is necessary to find possible restrictions made due to the data in an early stage of the project. Because only people who have expert knowledge in the underlying data can answer questions, the data must be analysed in an early stage of the project to address the questions to the experts.

The goal of this analysis is to get to know the available secondary data for this research project and to answer the first research question:

- What are possible actions, which can be suggested in smart homes to lower energy usage?

This question could be answered in a very general way, without any consideration of the underlying data for this project. Actions, which can be suggested to lower energy consumption in a household, could reach from changing traditional light bulbs to more energy efficient products, to suggestions to close the windows when the room temperature drops too much during ventilation. However, since the aim of the project is to make recommendations based on consumer behaviour data, the actions are derived from the smart home event data provided by Aizo.

The secondary data provided for this project was collected in alpha and beta test houses using the proprietary system digitalSTROM from Aizo. Since the approach developed in this project was designed to work for home automation systems, which record user behaviour data, in general, the digitalSTROM system will be referred to as “smart home system” in the further course of this document.

4.2 STRUCTURE OF DATA

The data provided from the smart homes is parsed from the log files of the system and is structured according to the entity relationship diagram in Figure 7.

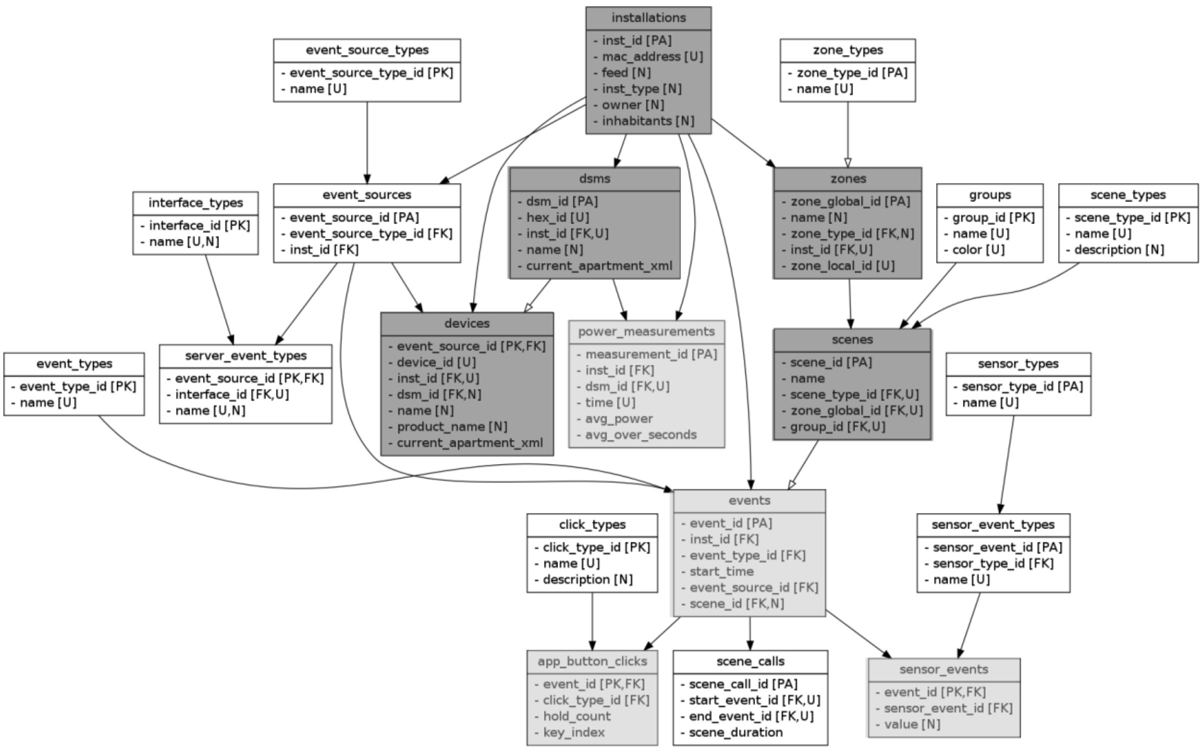


Figure 7 Entity Relationship Diagram (ERD) of the underlying data

The basic structure of the core entities (dark grey) can be explained hierarchically. The root node is represented by an installation, which describes one household (or in exceptional cases one house unit). An installation has several DSMs (abbr. for "DigitalSTROM Meter"), which represents one electric circuit in a smart home, and several zones (rooms). These zones are categorized in zone types and are related to one or many scenes, which again can be clustered again in groups and scene_types. The DSM is subdivided in the devices that are connected to the circuit. The dynamic entities (light grey) contain timelines for the core entities. Power_measurements are recorded per DMS and events are related to devices (event_sources) and depending on its type scenes to app_button_clicks or sensor_events. Event sources are again categorized in event_type_sources and server_event_types. In addition, app_button_clicks and sensor_events are characterized by type (click_type respectively sensor_type).

Because the data provided is parsed from log files, its structure is not fully normalized. The current dataset is just an intermediate step to enable further research on the data. The data points, which are relevant for the system developed in this project, will be covered in detail later.

4.3 EVALUATION OF THE VARIABLES OF ANALYSIS

The two variables of measurement, which are important for this research project, are "power measurements" and "events". While the other entities rather represent static information about the smart home system installed in the house, the two entities contain dynamic timelines, which can be used to deduce human behaviour patterns. A simplified version of the structure of the data is shown in Figure 8.

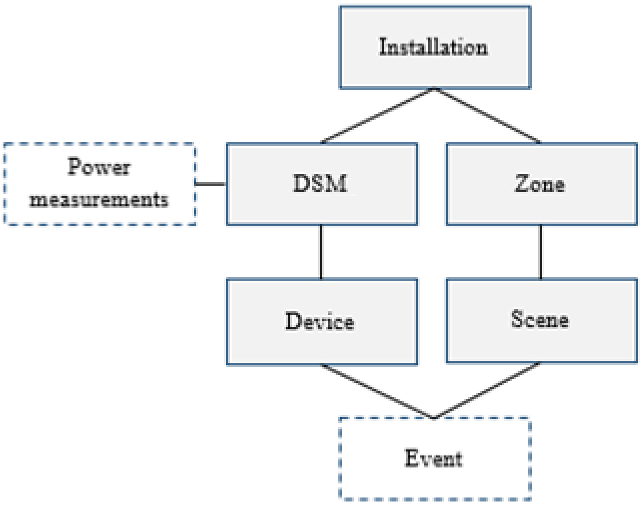


Figure 8 Power measurements and events in the structure of smart home data

To find possible actions to lower energy usage in smart homes, it is important to identify figures, which explain the energy usage in a smart home as detailed as possible. The most important attributes for a sequential smart home data event were defined as followed.

- Which device was triggered?
- When was the device triggered?
- What was the source of the event?
- How much power was consumed before?
- How much power was consumed after?

However, the data provided does not allow evaluating all these attributes, neither by the use of power measurements nor by events. The problem of the power measurement data lies in its structure. Power measurements can only be evaluated per DSM, but not per single device. Due to this fact, the power measurement data captured is only a summary of the energy consumed in the overall electric circuit from all devices together in this DMS, which makes it impossible to calculate the consumption of a single device.

Furthermore, the matching between the timestamps of the event and power measurement data resulted in no quantifiable results. The timestamps of the power measurements are stored in milliseconds, whereas the accuracy of the event data is only seconds. In addition, both values can introduce latency, which makes a reliable matching of the two entities impossible.

However, the attributes from the event data are well suited for identifying the attributes proposed. Since they are referenced to the device and to the scene table, both, the triggering device and the action triggered by the event can be identified. Nevertheless, there is no attribute to quantify the power usage, to measure the increase or decrease of the electric energy consumption by the given event. However, to answer the hypothesis defined for this project, the amount of energy saved is no critical factor, it is just important to identify if an event did increase or decrease the power usage (on or off event) which can be identified for the event data (see chap. 4.6).

Because of the arguments presented in this subchapter, the variable of analysis for this project will be defined as the smart home events.

4.4 DEFINITION OF AN EVENT

Based on the structure of the data and the variable of analysis, an event in the context of this research project can be defined as follows.

A single action caused by an inhabitant, which was recorded by the smart home system.

Obviously, there are more actions caused by an inhabitant, which are not recorded by the system. They cannot be considered as events, because they do not exist in the dataset provided. Events recorded by the smart home system can be divided in five different types: Scene calls, Sensor events, App button clicks, Local scene calls and Undo scenes, ordered descending by the number of occurrences in the data (Figure 9).

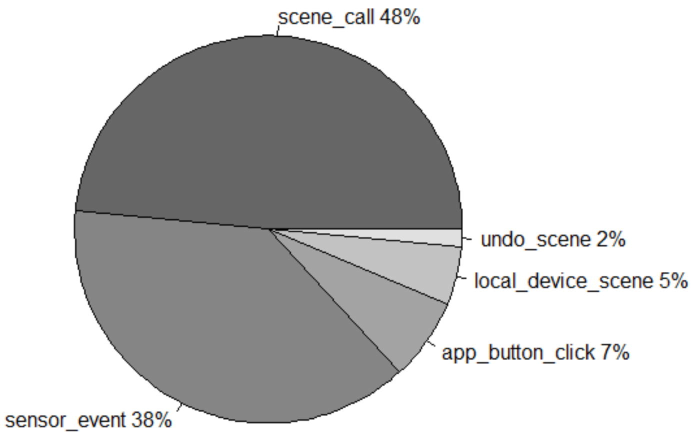


Figure 9 Percentage of events per event type

Because of their different characteristics, each event type has different attributes. Therefore, the relevant information for each event type is stored in separate tables in the database. However, to mine patterns in these events and apply the data for the recommender system, the events must be unified to one single timeline with identical attributes. The relevant attributes for the system are:

- event_id: Unique identifier for each event
- start_time: The exact timestamp, when the event occurred in the given smart home
- event_source: Which device of the smart home system did record the event
- event_name: What is the purpose of the event

These attributes lay the basis for the following steps and are important for the pattern mining process. The event_id will serve as unique identifier for each instance of an event and the start time allows ordering the events correctly in the event stream. To identify the relevant patterns, the source and the name will be important (see chap. 5.3).

The characteristics and the relevance for the research project for each event type is elaborated in the following subchapters.

4.4.1 Scene call events

A scene or also called “ambience” is a predefined configuration for one or more devices controlled by the smart home system. A simple example for a scene call is to turn on a spotlight in the living room whereas a more complex example is a “good morning scene” that includes several lamps and other devices like the cafe machine, which are turned on by calling this scene.

The data provided does not reveal which exact devices are called by a scene. However, the device referenced to a scene call is the event source (e.g. the switch, which was pressed by the inhabitant to call this scene). Therefore, one scene call is reported as one event in the database independent from the fact, how many exact devices where operated.

The system offers predefined scenes, but inhabitants can create new scenes and name them individually. For most of the scenes, the name attribute gives more information about its purpose. A scene call is always a direct interaction of the user with the system. The scene call properties are mapped as followed:

- event_source: The switch, which was the origin of the scene call (device.name)
- event_name: The name of the scene called (scene.name)

4.4.2 Sensor events

Sensor events are indirect interactions of the user with the system. The inhabitant is not actively using a device of the smart home system, but a sensor recorded that the state inside the smart home changed. Examples for sensor events are motion-detection events or the detection that the coffee machine was turned off. A sensor event consists of two major properties, which can be mapped as followed:

- event_source: The sensor, which reported the event (device.name)
- event_name: The value reported by the sensor (sensor_event_type.name)

4.4.3 App button click event

Besides the traditional functionalities, this smart home system offers also the possibility to control certain functions by using a mobile device. These interactions from the inhabitants with the system are represented by app button click events. The events can be mapped to the other events as followed:

- event_source: Which device was triggered from the mobile device (device.name)
- click Type: Type of click executed (click_type.descripition)

4.4.4 Local scene call

Besides the possibility of calling a device together with other devices in a scene, a device can also be called locally. These event types are especially used for devices with built in on/off switches, such as floor lamps, which is not directly connected to the system. If the state of such a local device has changed, the smart home system records this event as local scene call.

- Event_source: Name of the local device which was triggered (device.name)
- Event_name: The name of de scene called (scene.name)

4.4.5 Undo scene

A very small proportion of all events are classified as undo scenes. They are used to undo scene calls made before. One example to demonstrate the need of an “undo scene” is the “panic scene”, a predefined scene by the system. A panic scene call turns on all lights connected to the smart home and all shutters are lifted. To restore the initial status of the system, the inhabitant can do this for each device affected individually or by calling the undo scene.

Because an undo scene is only triggered in exceptional cases and is not helpful to identify behaviour patterns in a smart home, it will not be considered for this project.

4.4.6 Underlying data

Because of the reasons mentioned in this chapter, the event data set for this research project is defined by the following event types:

- Scene call events
- Sensor events
- App button click events
- Local scene call events

Furthermore, all entities with references to the event tables are part of the data as shown in Figure 10. This scope will be referred to as "underlying data" in following chapters.

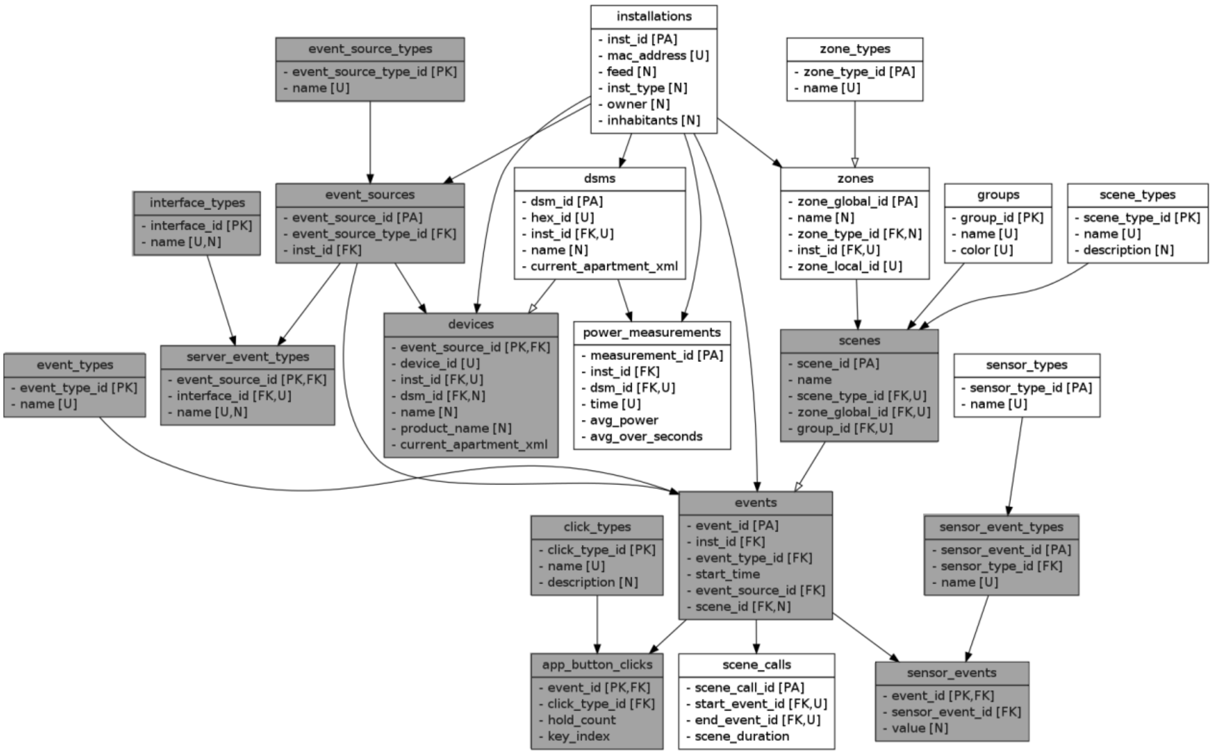


Figure 10 Underlying data highlighted in the ERD of the data provided

4.5 ANALYSIS AND VISUALISATION OF THE DATA

While the first subchapters of this fourth chapter defined the structure and scope of the underlying data of this project, the aim of this subchapter is to get an impression of the amount and content of the data.

The historical dataset provided by Aizo contains 33 installations with 3'521 devices, which are related to 4'331'443 events and 6'829 scenes. These events extend over a period between 8.12.2002 and 25.6.2014.

A closer analysis and especially the visualisation of the dataset revealed that the years before 2012 were only populated with few events that are very distant from the others. A real timeline was not apparent before. Therefore this events where identified as outliers. To ensure that no outliers are distorting the result of this study, the aim of the first part of the analysis and visualisation of the data was to find and remove such outliers from the historical dataset.

4.5.1 Outlier detection and removal

To detect and remove outliers in the dataset, the underlying data was grouped and visualized by all its attributes. Points that were very distant from all other observations were analysed in detail. As example, the outliers in the time dimension are presented. In Figure 11 all events from the underlying data are grouped by day, which reveals the outliers before the year 2012.

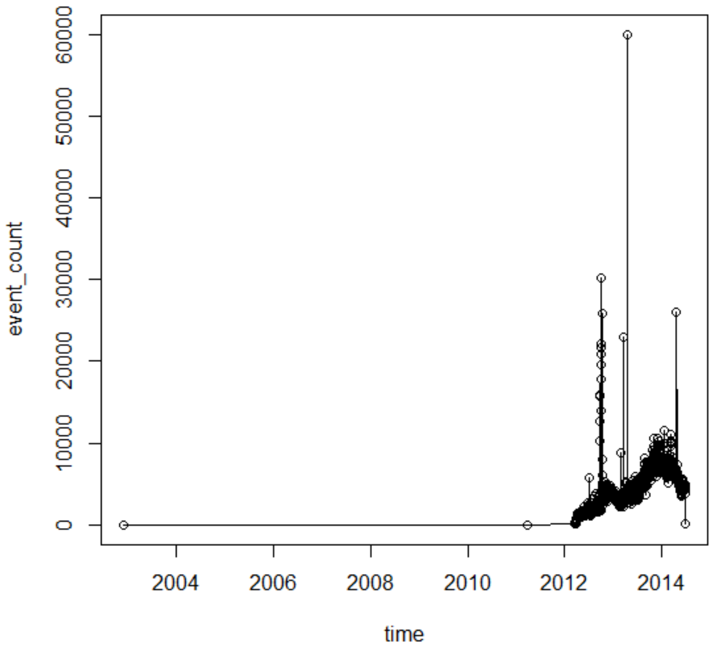


Figure 11 Outliers in boxplot of events grouped by date

Since the experts from AIZO confirmed, that the dataset should not contain data before 2012, the outliers were removed. However, a visualisation of the remaining events (Figure 12) showed that there were still outliers, which had to be analysed and removed individually.

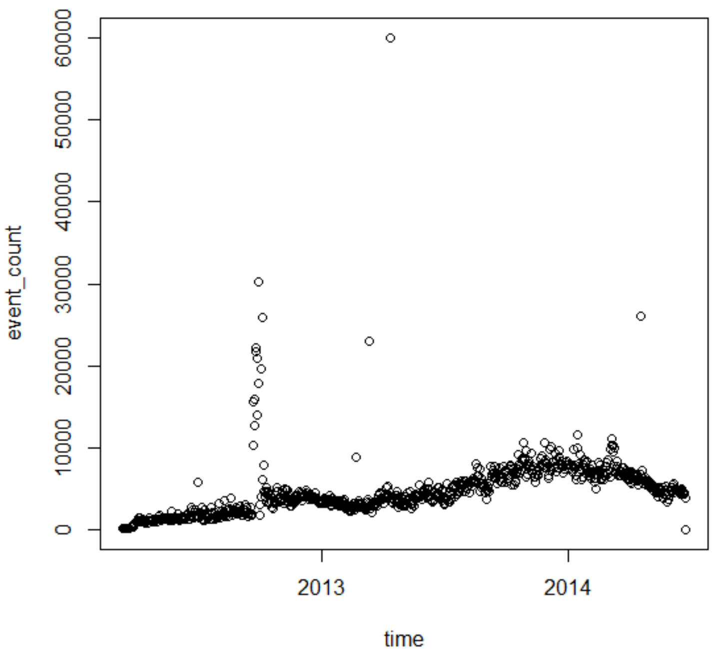


Figure 12 Distribution of the events after the removal of the outlier before 2012

By the visualisation of all other attributes, further outliers where found and removed from the dataset (described in appendix: Outlier removal description). All outliers could be assigned to a malfunction of a device (producing numerous events within seconds) or to tests conducted in the smart homes.

Although all outliers where removed, the data showed still large discrepancies between the installations. The total events per installation differ from 1’761 to 637’570. The shortest time series of events for an installation is only 20 days long whereas the longest series contains over 800 days of data. These numbers are summarized in the following Table 4 and visualized in the following figures, Figure 13 and Figure 14.

inst_id	oldest event	newest event	total events
1	4/23/2012	6/24/2014	42466
2	7/4/2012	6/25/2014	71621
3	4/1/2012	6/25/2014	227485
6	3/31/2012	6/25/2014	113015
7	5/12/2012	6/24/2014	178346
16	3/29/2012	5/31/2014	159177
31	3/22/2013	6/24/2014	97563
32	4/17/2012	6/24/2014	95738
35	3/24/2012	6/24/2014	144641
40	6/13/2013	6/24/2014	17636
49	5/4/2012	4/21/2014	637570
51	3/16/2012	6/24/2014	40461
52	4/13/2013	6/24/2014	37121
69	3/19/2012	3/22/2014	46629
71	5/24/2013	6/24/2014	39577
73	3/8/2013	6/24/2014	103820
79	4/3/2012	6/24/2014	38803
83	3/28/2012	6/24/2014	42012
84	3/25/2012	6/24/2014	329884
90	3/30/2012	6/24/2014	44371
92	6/8/2013	5/8/2014	56883
95	3/13/2012	4/27/2014	240735
96	3/21/2012	3/22/2014	77713
97	7/8/2012	3/22/2014	67415
102	7/23/2013	6/24/2014	30933
104	7/19/2013	6/24/2014	14236
105	8/6/2013	6/24/2014	84700
109	8/27/2013	6/24/2014	44174
114	9/3/2013	6/24/2014	25575
123	10/15/2013	6/25/2014	52237
124	10/17/2013	6/24/2014	102888
128	11/8/2013	6/24/2014	22537
190	6/4/2014	6/24/2014	1761

Table 4 Newest, oldest and total events per installation

The summary of these figures provides an overview of the data and was therefore used as a pre-selection for the installations considered for an evaluation of the system. At first, installations with longer timelines are considered as more interesting for pattern mining. Furthermore, all installations, where the newest event is not up-to-date, are no longer considered for the evaluation. The following Figure 13 gives an overview of the timelines of all installations.

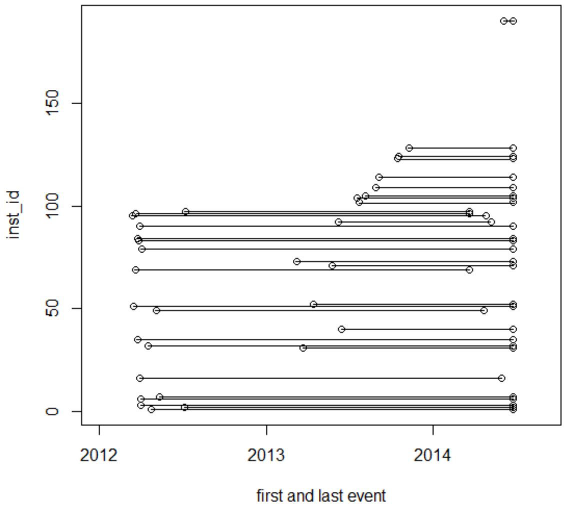


Figure 13 Timeline of all installations per installation id

In addition, the total number of events differs between the installations. The scatter plot and the histogram of the total events per installations (Figure 14) give an overview of the distribution of the events per household.

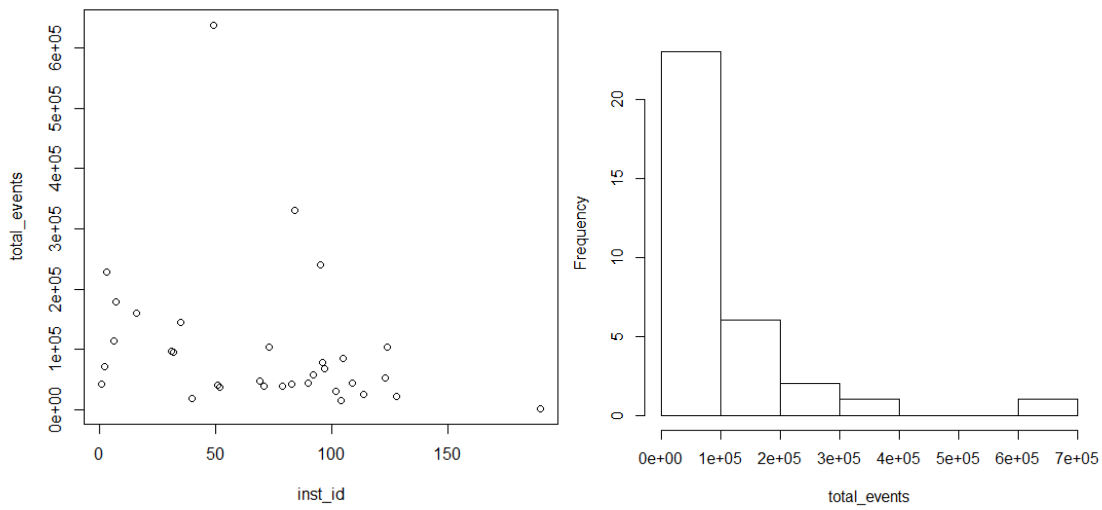


Figure 14 Distribution of the total events per household

4.6 POSSIBLE ACTIONS TO LOWER ENERGY USAGE

To identify possible actions to lower energy usage in smart homes, different approaches were taken into account. All actions that directly succeed in a decrease of energy usage were considered for a close analysis. These approaches are presented in this subchapter.

4.6.1 General approach

A “possible action to lower energy usage” in smart homes (in the following course of the paper referred only as “action”) is characterized by the fact, that after the execution of this action, the energy consumption of the concerned device(s) is lower than before the execution. This can be achieved either by turning them off completely or by reducing its consumption. Monetary aspects such as the amount of energy saved are not part of this study.

However, the most effective and most obvious approach to lower energy usage in smart homes is to turn all running devices off. Nevertheless, in order to fulfil the second part of the hypothesis, “not decrease comfort levels”, the actions must meet the interests of the inhabitants.

Summarized it can be said that an action is characterized by two main attributes.

- lower energy usage
- not decrease comfort levels

4.6.2 Actions in the underlying data

Actions can be seen as a subset of all events in in the underlying data. The challenge for an autonomously acting recommender system is to differentiate between events and actions. To determine actions from the underlying data, clear criteria must be defined, to distinguish them from normal events. The primary aim of the criteria is to identify events that fulfil the first attribute (lower energy usage). In a second part, the system should only suggest actions that also fulfil the second attribute (not decrease comfort levels).

Actions can be found in scene call events and in sensor events. The following subchapters will analyse both event types in detail.

4.6.3 Scene call event actions

A scene is characterized by the attributes Name, SceneType, Zone, and Group. The group attribute classifies the scene in one of the nine categories displayed in Table 5.

group_id	name	total events
0	Broadcast	295122
1	Licht	1223000
2	Schatten	99513
3	Klima	9081
4	Audio	13219
5	Video	12464
6	Sicherheit	3
7	Zugang	24
8	Joker	0

Table 5 Scenes are classified in nine groups

The analysis of the scene group exposed that a majority of events that occur in a smart house are classified as “light” (Figure 15). Therefore, it is not surprising, that most of the actions, which will be presented at the end of the chapter, will be connected to light. However, the attempt to use the group attribute as a reference for the decision, if an event can be used as action, did not lead to the desired results.

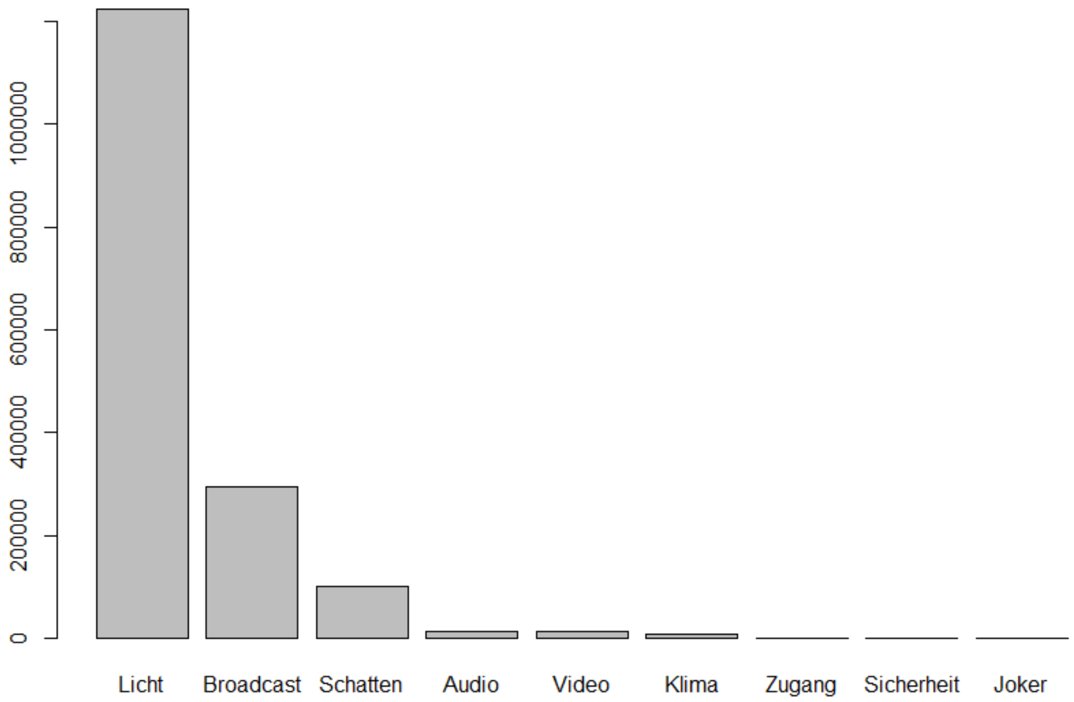


Figure 15 Total of events per scene group

Furthermore, also the zone attribute gives no indication about the usefulness of an event to be used as an action. However, by analysing the name attributes of the scenes in the underlying data manually, it became apparent, that the name of a scene does reveal certain information about its purpose. A simple rule was defined: If the name of a scene contains one of the following words, it is considered as an energy saving event and therefore as an action for the system:

- Absent (a predefined scene which turns off all devices when leaving the house)
- Dim (scenes to dim devices)
- Off (scenes to turn off devices)
- Sleep (scenes to turn off devices before going to sleep)
- Standby
- Stop

In the underlying data, the execution of this rule leads to an overall list of 20 scenes, which are called by 1’283’756 unique events, shown in Table 6.

scene name	event count
Absent	8964
Dimm-Area	417
Off	907269
Off-Area1	28936
Off-Area2	27040
Off-Area3	6221
Off-Area4	2970
Off-Device	113963
Off-Stimmung10	10146
Off-Stimmung20	505
Off-Stimmung30	12
Off-Stimmung40	70

Sleeping	2712
Stop	25369
Stop-Area1	979
Stop-Area2	768
Stop-Area3	543
Stop-Area4	474
Zone-Deep-Off	102736
Zone-Standby	43650

Table 6 Scene events that can be considered as actions

Experts from Aizo confirmed that the scenes from Table 6 can be considered as actions, because they lower energy usage.

4.6.4 Sensor event actions

Sensor events are defined by sensor event types, which are described by a name attribute. Similar to the scene events, the name attribute reveals the purpose of an event and a rule was created to distinguish the actions from the regular sensor events. If the name of the sensor event does contain one of the following words, it can be considered as action:

- Aus (German for "off")
- Off
- Uit (Dutch for "off")

In the underlying data, the execution of this rule results in an overall list of 55 unique sensor event type names, which are referenced from 143'030 events

name	**count**
Fernsehen aus	676
1OG Till TV aus	166
Brenner aus	726
Dampfabzug aus	1233
Fernsehen aus	122
Fernsehen ausgeschaltet	538
Fernseher aus	1262
Gefrierschrank aus	2123
Gerät aus	43
Hz Brenner aus	5608
Hz Pumpe Fussboden aus	3
Hz Pumpe Warmwasser aus	116
Kaffee1000Waus	2092
Kaffee200Waus	108
KG_Bespr Wasserkocher aus	11
Kleiner Backofen aus	51
Krups Nespresso CitiZ & Milk aus	14365
Kuehlschrank aus	25957
Kuehlschrank unten aus	11020
Kühlschrank aus	2755
Kühlschrank Garage aus	1484
Kühlschrank ZWS aus	3819

Luftbefeuchter Schlafzimmer aus	102
Mikrowelle aus	299
Mr. Tea Samowar aus	237
Nespresso Kaffee einlassen aus	124
Nespresso U Zwischenstecker aus	24
PC turn off	245
Radio aus	112
Samsung TV aus	366
SC TV&Sat aus	28
SC TVSat aus	1011
Spuelmaschine aus	2616
Stecker TV aus	3195
Stromer Laden aus	21
Teekocher aus	2985
Till Video aus	109
Toaster aus	1538
TV aus	227
TV uit	461
Ventilator aus	2
Ventilator Bad turn off	610
Verbrauch aus	93
Video Geräte uit	2344
Wäschetrockner aus	7
Waschmaschine aus	47121
WaschmaschineAus	483
Wasserkocher aus	369
WasserkocherAus	1359
WZ TV aus	2148
ZWS Carotec aus	258
ZWS Fernsehen aus	236
ZWS Fernseher aus	7
ZWS Test Verbrauchsmeldung aus	13
ZWS Wasserkocher aus	2

Table 7 Sensor events that can be considered as actions

Experts from AIZO confirmed, that also the events in Table 7 can be considered as actions, because they lower energy usage.

4.7 SUMMARY OF CHAPTER 4

In this chapter, the data provided for this research project was analysed with the aim to find possible actions that can be suggested, to lower energy consumption in smart homes and to answer the corresponding research question 1.

At first, the structure and the content of the data provided was analysed, to identify the sensors and devices connected to the smart home system. The apposition of additional data or data sources is not preferred, because this would result in manual effort of the inhabitant, which would have a negative impact in the comfort of the user. Therefore, the data set provided is the single source of information for the recommender system and the system can only actions that are already known by the system.

A closer analysis of the data revealed, that the power measurement data is not recorded as precise as necessary for obtaining information about the power consumption of a single device. In contrast, the event data is well suited and therefore eligible to find possible actions. Interviews with the experts of Aizo were subjected, to identify the relevant events out of the event data, which did lead to the final scope of data, the underlying data.

To identify the actions from the underlying data, rules where defined, to distinguish actions from normal events. Due to this fact, these events can be suggested as actions to lower energy. The rules defined for this project are well adapted for the data provided by the smart home automation system from Aizo. For other datasets in other smart houses, the rules might look differently. However, summarized it can be said that possible actions, which can be suggested in smart homes to lower energy usage, are special occurrences of regular events. It is essential to identify attributes, which allow rules to distinguish such actions from normal events.

CHAPTER 5 - RELEVANT BEHAVIOUR PATTERN FILTERING

5.1 INTRODUCTION OF CHAPTER 5

The following chapter takes place in the suggestion phase of this master thesis and is going to identify relevant behaviour patterns in the underlying smart home event data. Furthermore, it will address and answer the research question RQ2:

- Which types of behaviour patterns are needed to suggest actions?

In the previous chapter, actions were identified, which lower energy usage in smart homes. By suggesting these actions in smart homes, the first part of the thesis statement “It is possible to design a recommender system that can suggest actions in smart homes based on consumer behaviour, which will lower energy usage” can be fulfilled. However to realize also the second part of the hypothesis, which states that the recommendations should “not decrease comfort levels”, the actions must be suggestion in the appropriate moment only. To predict the right moment to suggest these actions, relevant patterns are needed.

This chapter follows a top-down approach: At first, the characteristics of relevant patterns are identified and in a second step, it is exposed how these relevant patterns can be filtered of the result-set of a frequent (and/or periodic) pattern-mining algorithm. In the next chapter (chap.6), the most suited pattern-mining algorithm is evaluated to provide the input for the pattern filter, designed in this chapter (illustrated in Figure 16).

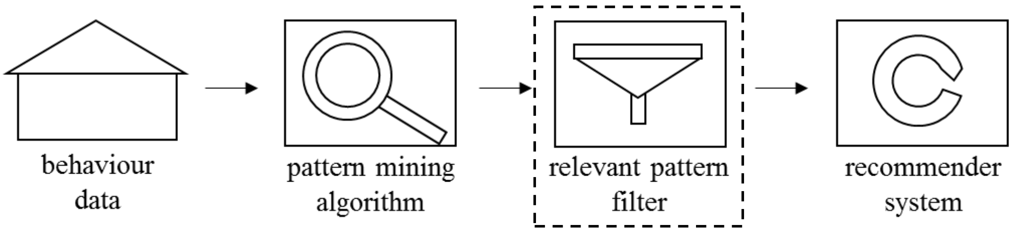


Figure 16 Relevant behaviour pattern filtering in the overall project

For a common understanding of the term “behaviour patterns”, the chapter starts with a definition of the term and of the attributes “frequent” and “periodic”.

5.2 DEFINITION OF BEHAVIOUR PATTERNS

The term “pattern” is often used in the literature without a precise definition, or at best with a definition postulated for the particular situation. To define a sequential (behaviour) **pattern** in the context of this research project, a definition published in the paper “Finding Sequential Patterns from Large Sequence Data” (Esmaeili & Gabor 2010) is used as basis:

> *“A pattern is defined as sequence of symbols which appears more than a threshold in a data sequence”* (Esmaeili & Gabor 2010).

Because the definition for a pattern of Esmaeili & Gabor covers only frequent pattern, the definition is stated more precisely. For this particular research project **a frequent and/or periodic (behaviour) pattern** is defined as follows:

> *A pattern is defined as sequence of events, which appear frequent and/or periodic in a data sequence.*

The term **frequent** is defined as follows:

Given a sequence database D = s1,s2,...,sn, the support of a sequence α is the number of sequences of D which contains α as a subsequence. If the support of α is bigger than a threshold maxsup, then α is a frequent sequence (Peng & Liao 2009)

According to the "optimal parameters" published in "Learning frequent and periodic usage patterns in smart homes"(Schweizer 2014), the threshold (maxsup) for frequent patterns in this project was defined to 0.005 for this research project.

A **periodic pattern** is defined as follows:

Periodic patterns are pattern, that" reoccur at always the same intervals" (Schweizer 2014).

According to the "optimal parameters" published in "Learning frequent and periodic usage patterns in smart homes"(Schweizer 2014), "in the same intervals" for this project is defined as followed: 75% of the intervals between the occurrences of the pattern are equally long (with a grace period of 10%). Whereas the grace period can defined as "percentage by which the interval between two consecutive occurrences of a pattern can vary" (Schweizer 2014).

5.3 RELEVANT TYPES OF BEHAVIOUR PATTERNS

This subchapter addresses the question (RQ2) about which types of behaviour patterns are needed to suggest actions. Because not all frequent or periodic patterns are relevant for the system, characteristics must be defined, to identify the relevant patterns in the result set of the frequent (and/or periodic) pattern-mining algorithm.

To ensure, that a relevant pattern can be used to suggest actions, it must be composed of two main components: At first, a relevant pattern must contain at least one action to lower energy usage in smart homes (action) as defined in chapter 4.6. Additionally, the pattern must consist of normal events, which serve as condition to suggest the action at the right time. Because a one-event condition is insufficient as evidence to suggest an action, the overall length of a relevant pattern must be at least three events (normally one action and two events, but in exceptional cases also three actions are possible, explained later in Figure 19). Therefore, a relevant pattern can be defined as follows:

A relevant pattern is a (behaviour) pattern that is longer than two events and contains at least one action to lower energy usage

Because a relevant pattern consists of normal events (condition) and an action, it can be interpreted as association rule. An association rule is an implication of the form:

$$X \rightarrow Y, \text{ where } X, Y \subset I, \text{ and } X \cap Y = \emptyset$$

The rule states that when X occurs, Y occurs with certain probability (Deenadayalan et al. 2014). This approach is illustrated in Figure 17.

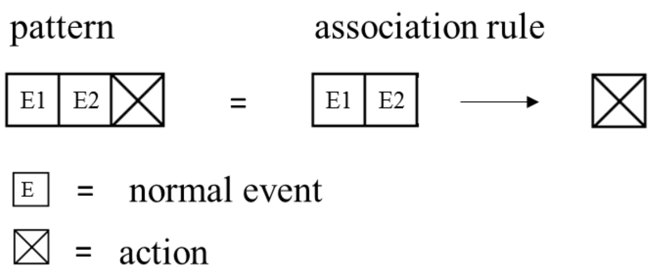


Figure 17 A relevant patterns can be interpreted as an association rule

Because association rules do typically not consider the order of the items, also patterns where the action is not at the end of the pattern can be interpreted as association rules. Under disregard of the time (which defines the order of the events in a pattern), the condition (normal events) of the rule does still conclude the action. Therefore, also patterns where the action constitutes the beginning or the centre are defined as relevant patterns (Figure 18).

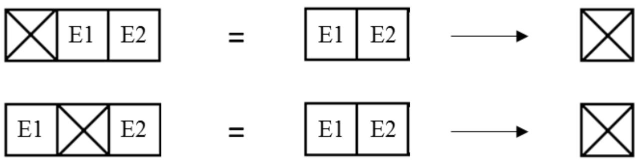


Figure 18 Patterns with an action in the beginning or in the middle

5.3.1 Patterns with multiple actions

As defined before, patterns can also contain two or more actions. To apply the association rule approach also for such patterns, one association rule is created for each action in a pattern. Because actions are just a special form of events, the other actions will be treated as normal events and serve as condition for the rule.

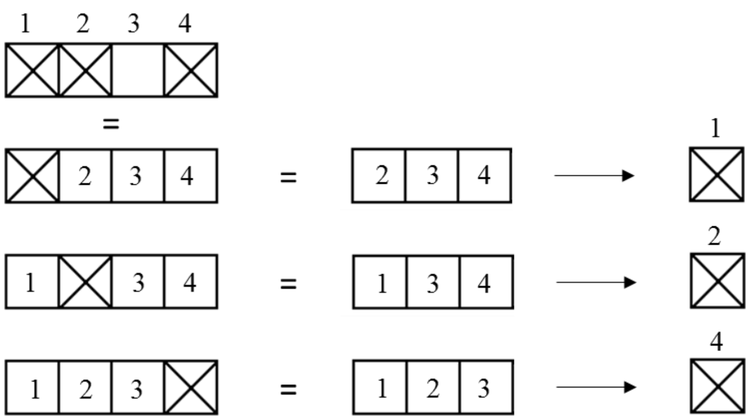


Figure 19 A rule is created for each action in the pattern

As illustrated in the example of Figure 19, a four-event pattern containing three actions will result in three rules. For each action (1, 2 and 4) in the pattern, a rule is created and the other actions are interpreted as normal events. They work as condition of the rule, in order to recommended the action if the condition appears in the behaviour data (without the action).

5.4 SUMMARY OF CHAPTER 5

In this chapter, the (research) questions about which types of behaviour patterns are needed to suggest actions is addressed and answered. Based on the analysis and the definition of the underlying data, (behaviour) patterns in the context of this research project were defined. Furthermore it was derived which requirements must be met, that a pattern can be considered as relevant. Simplified it can be stated, that types of behaviour patterns which are needed to recommend actions, must fulfil the following criteria.

- The pattern must occur frequent and or periodic in the data:
    - To be considered as frequent, pattern must reach a minimum support of 0.005.
    - To be considered as periodic, 75% of the intervals between the occurrences of the pattern must be equally long (with a grace period of 10%).

- The pattern must be relevant for action prediction:
    - To be considered as relevant they must contain at least one action to lower energy usage.
    - Additionally the pattern must have a minimum length of three events (or actions).

Due to these facts, it became clear which types of behaviour patterns are the input (frequent and or periodic patterns) as well as the output (relevant pattern) for the relevant behaviour pattern-filtering component, introduced in this chapter.

Furthermore, this chapter induced an approach to transform relevant patterns into association rules. This step is essential, that the recommender system introduced in the next chapter (chap. 6) can recommend actions based on the relevant patterns. Because the translation from patterns to rules has also impact on the criteria for relevant patterns, the approach is introduced in this chapter.

CHAPTER 6 - ALGORITHMS TO FIND RELEVANT PATTERNS

6.1 INTRODUCTION OF CHAPTER 5

The following chapter takes place in the suggestion phase of this master thesis and is going to introduce the mining of relevant patterns in the underlying smart home event data. Therefore, it will address and answer the research question RQ3:

- What algorithms can be used to find relevant patterns in the behaviour data?

Since a lot of research was already conducted in the field of mining frequent (and/or periodic) patterns in sequential (event) data (Srikant & Agrawal 1996; Esmaeili & Gabor 2010; Li et al. 2010; Rashidi & Cook 2013; Schweizer 2014), the development of such an algorithm is not scope of this project. Nevertheless, such an algorithm is essential for the development of the recommender system in this project and therefore the most suited (existing) algorithm will be evaluated.

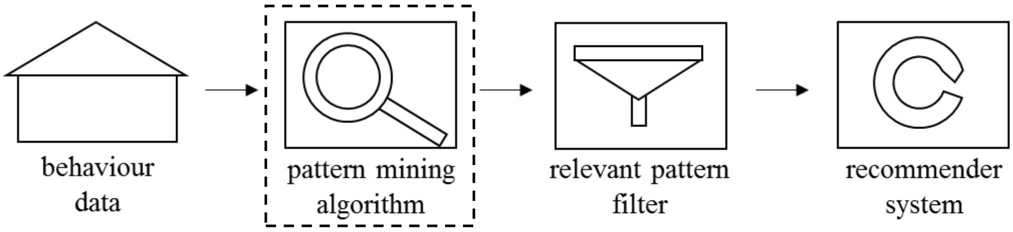


Figure 20 Pattern mining algorithm in the overall project

Based on the findings of the chapters “Data Analysis” (chap. 4) and “Relevant behaviour pattern filtering” (chap. 5), the input and the expected output of the pattern mining algorithm are already defined (illustrated in Figure 20). In the first part of the chapter, the most suitable pattern-mining algorithm is evaluated. Because the algorithm closes the gap between the underlying behaviour data and the relevant pattern filter, (relevant) sample pattern can be mined after the evaluation and adaption of the algorithm. These sample patterns are presented in the second part of this chapter (chap 6.3).

6.2 EVALUATION OF ALGORITHMS

In this subchapter, algorithms are evaluated, to mine frequent and periodic patterns in sequential smart home event data, as defined in the previous chapter (chap. 5). Because a proper and comprehensive evaluation of such algorithms for the usage in smart homes has been published recently (Schweizer 2014), the selection of the most appropriate algorithm for this research project will be based on these results.

The three established frequent sequential pattern-mining algorithms PrefixSpan, BIDE+ and GapBIDE as well as their adaptations were evaluated by Schweizer (2014) in his paper “Learning frequent and periodic usage patterns in smart homes” for the usage on smart home event data. They were benchmarked against each other and against a self-developed algorithm with the name WSDD. All four algorithms mined a reasonable amount of frequent sequential patterns for the input parameter set of pattern length = 2-7 events, minimum support = 0.01 – 0.001, overlapping patterns and wildcarding deactivated.

The traditional frequent sequential pattern mining algorithms like PrefixSpan, BIDE+ or GapBIDE require pre- and post-processing to be used for mining smart home event data. Furthermore, if different minimum and maximum lengths of patterns shall be mined, those algorithms need to run multiple times to report the correct support count. Moreover the study exposed that wildcarding could not fulfil the expected potential because no significantly higher support counts were mined with wildcarding being activated. Wildcarding will therefore not be considered for this project.

The run times of the different algorithms showed large deviations, while BIDE+ needed the longest to mine the same patterns as the other three algorithms. Both, GapBIDE and especially PrefixSpan run significantly faster. However, they all were outperformed by the algorithm WSDD (Figure 21).

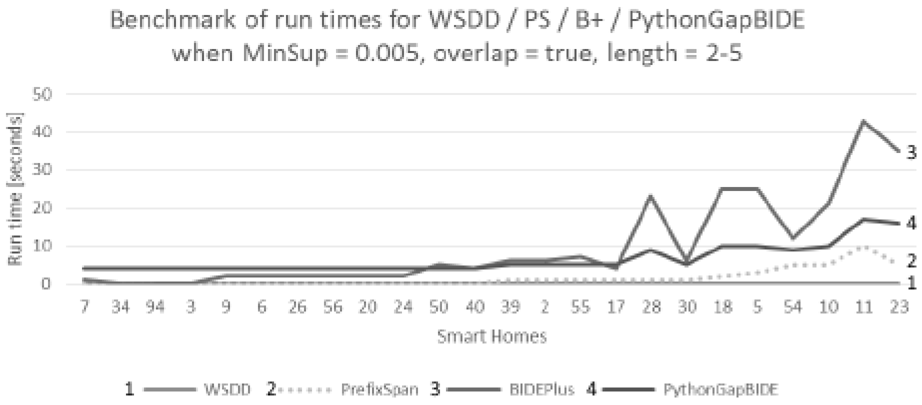


Figure 21 Run time comparison of the algorithms evaluated by Schweizer (2014)

Although all of the evaluated algorithms performed relatively poor in terms of memory consumption, the evaluation revealed a clear winner: "WSDD, […] is an adequate choice for sequential pattern mining in the smart home area" (Schweizer 2014). The advantages of WSDD over other algorithms for mining patterns in sequential smart home data can be summarized as follows:

- It is the fastest generic sequential pattern mining algorithm, which was benchmarked in the publication of Schweizer (2014).
- WSDD needs no pre-processing for the input data. While generic sequential pattern mining algorithms like BIDE+, PrefixSpan or GapBIDE, need a sequence database as input and therefore a pre-processing step, the continuous event data from the smart homes can directly be fed to WSDD.
- In contrast to other algorithms, WSDD is able to report the correct support (count) for each pattern. BIDE+, PrefixSpan and GapBIDE only report the correct support, when using multiple runs for the same smart home.
- WSDD offers the possibility to mine periodic patterns.

Because of the findings of Schweizer (2014), the question for the most suited frequent and periodic pattern mining algorithm for this project can be answered. The algorithms BIDE+, PrefixSpan and GapBIDE were all successfully tested to mine frequent patterns in sequential smart home event data and could therefore be used to find relevant patterns in this research project. However, because of the advantages mentioned above, WSDD is selected as foundation for the recommender system developed in this master thesis. Whereas the differences in speed are no critical factor in this project, the three other points mentioned above, are difficult to renounce for the recommender system developed in this project.

Therefore, the research question (RQ3) about algorithms that can be used to find relevant patterns in the behaviour data, can be answered. The combination of the pattern-mining algorithm WSDD and the relevant pattern filter designed in the previous chapter is the most suited algorithm to find relevant pattern in the underlying smart home behaviour data. This combination will be referred as "relevant pattern-mining algorithm" in the further course of this document (Figure 22).

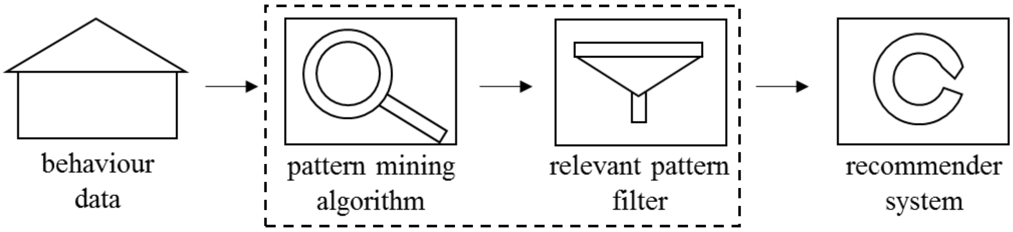


Figure 22 Relevant pattern-mining algorithm to find relevant patterns in the overall project

6.3 SAMPLE PATTERNS

In this subchapter sample pattern from the underlying data are presented, which were mined by the relevant pattern-mining algorithm. An analysis of the sample pattern exposed, that the criteria for the definition of "relevant patterns" stated in this chapter (chap. 5.3) must be adapted slightly to conform to the context of the underlying data. The reason for this adaption is that still patterns where mined, which are not useful to suggest action in smart homes and therefore are not considered as relevant. Examples for relevant patterns and exceptions mined in the underlying data will be presented in the following two subchapters.

6.3.1 Relevant sample patterns

The relevant patterns mined in the underlying data showed lengths of 3 to 6 events. Whereas the minimum of three events is a predefined parameter, which was discharged by the definition of relevant (chap. 5.3), the maximum was not defined. However, the testing on the underlying data showed that no relevant pattern was longer than six events.

The following table presents three examples of the relevant patterns mined in the underlying event data by the relevant pattern-mining algorithm. The grey background of the cell indicates the action, which must be part of a relevant pattern (Table 8).

P.	Event 1	Event 2	Event 3
1	Luz_Cozinha_+_Interruptor:Off	Luz_Escada_Keller_+_Interruptor_EG: Stimmung1	BM_Garagem_Corredor: State_1
2	Taster_Waschküche:Stimmung1	Taster_Waschküche:Off	Taster_Abgang_Keller:Off-Stimmung10
3	Taster_Abgang_Keller:On-Stimmung11	Lampe_Steinkeller:On-Device	Lampe_Steinkeller:Off-Device

Table 8 Relevant sample patterns mined in the underlying data

6.3.2 Exceptions

The mining of sample patterns with the relevant pattern mining algorithm produced also patterns, which do fulfil the formal requirements for relevant patterns (chap. 5.3), but are still not useful to make recommendations for energy saving, because of their smart home automation system specific character. These exceptions might be different for each smart home system, because of different sensors, scripts or other system-specific individualities and cannot be generalized. However, to make a field test of the system possible, these exceptions had to be included to the relevant pattern filter (chap. 5).

The smart home automation system provides the possibility for the inhabitant to implement and deploy their own automation scripts. Such scripts are used to enhance the comfort of the inhabitants. Some of the inhabitants use the scripts for the automation of well-known daily routines, which does interfere with the purpose of the system developed in this project. Certainly, such routines are found with a high support count in the frequent and periodic patterns. Furthermore, some of these routines are even considered as relevant, because the aim of the script could be to turn off a device at the end of the routine. Fortunately, such patterns can be identified by the "source" attribute of the action, which always refers as "script". These patterns are removed from the relevant patterns in the post-filtering step, because a recommendation to “script off” a device would be useful for nobody and lower the comfort of the inhabitants. Two examples of such “scripting” patterns are listed in Table 9.

P.	**event 1**	**event 2**	**event 3**
1	BM_Keller:State_0	BM_Garagem_Corredor:State_0	Scripting:Off-Area2
2	FL_Bewegungsmelder:State_1	FL_Bewegungsmelder:State_0	Scripting:Off

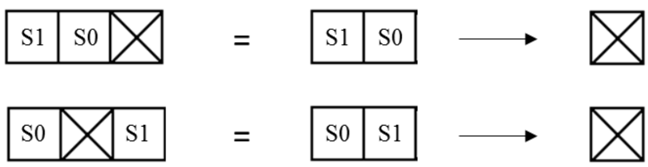
Table 9 Sample patterns with contain "scripting" actions

Although motion detection devices are no prerequisite of the system, some of the households, where the underlying data was collected do use such devices. Even though in many publications (Aztiria et al. 2011; Rashidi & Cook 2013; Rohini & RajKumar 2014), the positive effect of such devices for smart home automation systems has been disclosed, they must be treated as an exceptional case in this project. The reason for this special treatment is that a motion sensor creates two events on each motion detected (one event when the motion is detected and one event when no motion is detected anymore). Two sample patterns where motion detection is involved are listed in Table 10.

P.	**event 1**	**event 2**	**event 3**
1	Bewegungsmelder:State_1	Bewegungsmelder:State_0	Taster_Alessandro:Off
2	Bewegungsmelder:State_0	Taster_Alessandro:Off	Bewegungsmelder:State_1

Table 10 Patterns with consist of two motion-detection events and one action

By using the association rule approach, introduced the previous chapter (chap. 5.3), these patterns are transformed into rules as illustrated below (Figure 23).



S1 = Bewegungsmelder:State_1, S0 = Bewegungsmelder:State_0

Figure 23 Rules deduced from the sample patterns in Table 10

These association rules deduced from the patterns, will have as consequence, that every movement detected by the motion detection device will cause a suggestion, in this case “Taster_Alessandro:Off”. An analysis of the behaviour data from smart homes, where such motion detection devices are installed, exposed that such rules would generate an average of over 70 suggestions per household and day. Such an overload of recommendations would certainly decrease the comfort levels of the inhabitants, because only few of the recommendations might be useful. However not all patterns containing motion detection sensor events are removed. In longer patterns, motion detection events combined with other events are helpful and can contribute to suggest the correct action. Therefore, only patterns, which consist entirely of motion detection events and actions, will be removed in the post-filtering step.

The last exception discovered in the sample patterns mined in the underlying data, arises if a pattern contains two or more times the same event: If in a three-event pattern, an action is surrounded by two similar events, the condition of the rule deduced will consist of two similar events only. If a pattern includes the same action twice, the action suggested is part of the condition and of the action of the rule at the same time. Moreover, if one of the actions is at the end of the pattern, the occurrence of the action directly leads to the suggestion of the action. Because such recommendations are not useful, these patterns should not be considered as relevant.

All exceptions exemplified in this subchapter will be excluded from the set of relevant patterns for the evaluation of the prototype with this underlying data and will therefore not be considered for recommendations. However, for the sensor data of other smart home automation systems, the data specific exceptions might be different.

6.4 SUMMARY OF CHATPER 6

In this chapter, algorithms that can be used to find frequent (and/or periodic) patterns in the behaviour data where evaluated. Because of the preliminary work of Schweizer (2014), four suitable algorithms for frequent and periodic pattern mining in smart home event data where identified. The four algorithms (BIDE+, PrefixSpan, GapBIDE and WSDD) where analysed and the most suited (WSDD) for this project was chosen.

Because of the visualisation of sample pattern, exceptional cases where found. These exceptions are caused by the characteristics of the smart home automation system used to record this event data and are only relevant for the usage of this underlying data. The exceptions identified are

- Patterns with "scripted" actions.
- Patterns, which consist only of motion detection events and actions.
- Patterns, which includes the same action twice.
- Patterns, which consist of similar events (instances of the same event type) only.

Due to these facts, it became clear which types of behaviour patterns are needed to suggest actions and what algorithms can be used to find such patterns.

.

CHAPTER 7 - DESIGN OF THE RECOMMENDER SYSTEM

7.1 INTRODUCTION OF CHAPTER 6

The following chapter of this master thesis takes place in the development phase of the project and presents details about the design and the implementation of the recommender system. At first, the general design of a recommender system is described and in a second step, the specific implementation using the underlying data is made known. Furthermore, the chapter will address and answer the research question (RQ4):

- How can such a recommender system be designed?

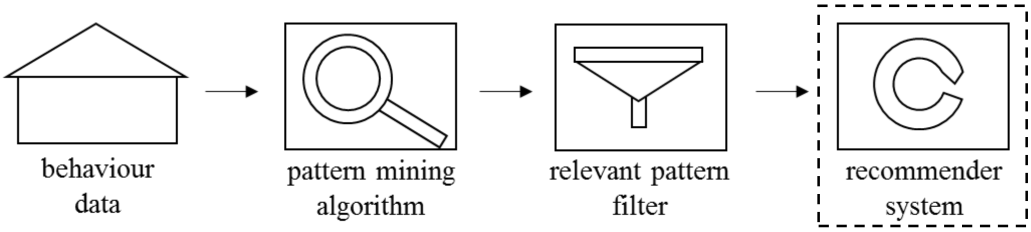


Figure 24 Recommender system in the overall project

7.2 ARCHITECTURE

In general, recommender systems are information filtering systems, which seek to predict user preferences. The problem of a recommender system is to suggest meaningful actions or items that may be of interest for the user of the system. The design of a recommendation engine depends very much on the domain and the particular characteristics of the data provided (Ricci, Rokach, Shapira, et al. 2011).

The data analysis of a recommender system can be based on a different variety of methods (Figure 25). The method chosen for this project is "Association Rules" (explained in chap. 5.3).

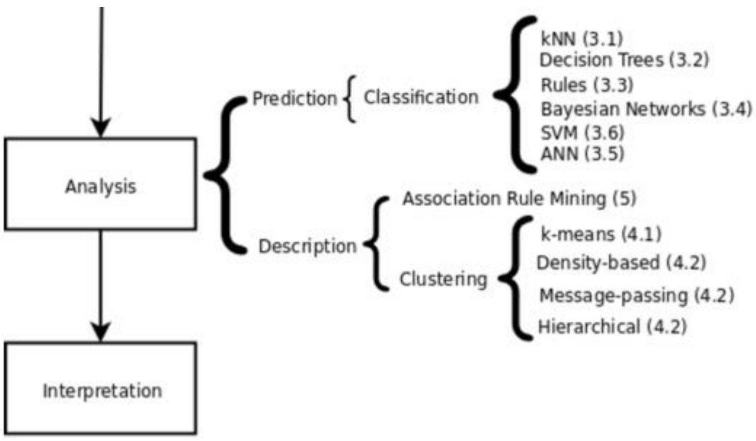


Figure 25 Data analysis methods in a recommender system (Ricci, Rokach, Shapira, et al. 2011)

The design of a recommendation engine depends very much on the domain and the particular characteristics of the underlying data. Although a system can be designed using different technologies and data analysis methods, it can be classified into one of the following three broad groups: Content-based systems, collaborative filtering systems, and association rules based systems (Ye 2011). In exceptional cases, these methods can also be combined. Because the system designed and implemented in this projects uses association rules for data interpretation, this approach can be attributed to "association rule based recommender systems".

The architecture of the recommender system developed in this project (illustrated in Figure 26) can be divided in tree main parts:

- The storage of the association rules.
- The event stream of the current behaviour data inside the smart home.
- The matching algorithm for both previous points.

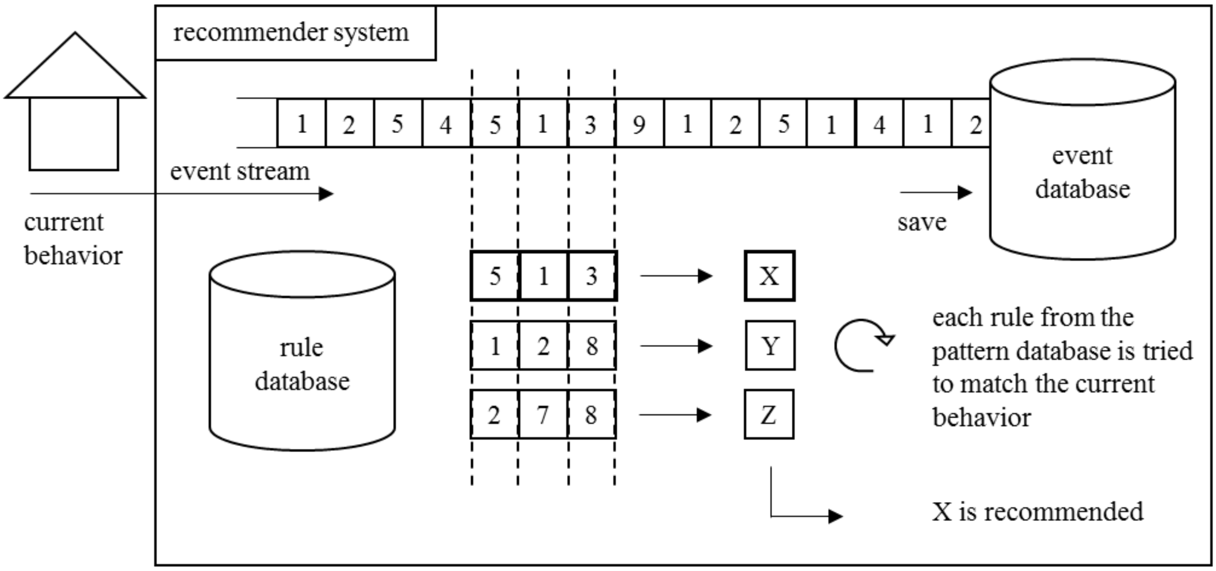


Figure 26 Architecture of the recommender system

The **rule database** stores the association rules, which were deduced from the relevant pattern.

The **event stream** contains the current events from the smart home, ordered by time of their occurrence. The fact that two events can occur at the very same time, increases the complexity of this problem. Either a second (unique) attribute is used to order two simultaneous events or the order becomes ambiguous.

The **matching algorithm** is the core component of the recommender system. It matches the rules and the event stream. The most common existing rule matching algorithm is RETE (Forgy 1982). The algorithm has been designed especially to enable efficient processing of large rule sets and is open source. Because the prototype implemented in this project contains a manageable amount of rules, the complex adaption of RETE was not reasonable. Instead, a more simplified concept of a deterministic finite state machines was implemented (Figure 27). Therefore, a new instance of the machine is created for each new event in the stream. However, if the matching did not work in the first attempt, the instance changes into state E (exit) and will be destroyed from memory. If the condition did match and the next event is not the action itself, the machine changes into state R and a recommendation will be sent.

The major disadvantage of this approach is the increased memory consumption because for each event, a new instance of a state machine per pattern will be created and must be held in memory. However, almost all of these instances will be destroyed after one event. Since memory consumption is no critical factor in this project, the simplified approach is chosen for this project.

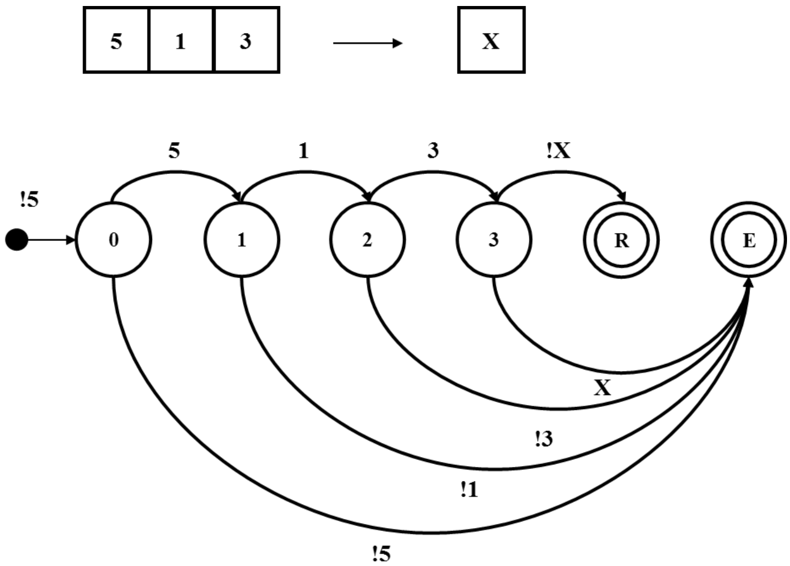


Figure 27 Example for the state machine created from an association rule

The example for the state machine in Figure 27 is for an association rule deduced from a pattern, where the action was at the end of a pattern. Therefore, the last condition before sending the recommendation is to check, if the next event is not the action itself, which should be recommended (!X). For other positions of the actions, the state machine will look slightly different: For patterns where the action is at the beginning, this check is required as first state of the machine and for patterns where the action is in the centre, the check is not necessary at all. As a consequence of this is, besides the rule also the position of the action must be stored.

7.3 PRIORITIZATION

The design of the recommender system allows, that more than one rule can match at the same time. For this reason, it is necessary that rules can be weighted, to decide which rule does match first. Furthermore, the prioritization criteria introduced in this chapter will also be used to exclude rules under a certain threshold, which is an extension to the relevant pattern filter introduced in chapter 5. This kind of filtering will be applied between evaluation phase 1 and 2 (chap. 8).

To calculate the weight of a rule, different indicators can be considered.

1) Support (count) of the pattern relative to all events
2) Number of events in a pattern
3) Position of the action in the pattern
4) Date when the rule was mined
5) Confidence of the pattern

While the support (count) is already calculated by the pattern-mining algorithm WSDD, confidence must be calculated after the mining by the recommender system. Compared to the other indicators (Number of events in a pattern, position of the action in the pattern, date when the rule was mined), the confidence of a rule is rather complex to calculate. Therefore, the formula is derived and exemplified in detail in the next subchapter.

7.3.1 Confidence

Confidence in the field of association rule mining is defined as follows:

> *The rule holds in T with confidence conf if conf% of transactions that contain X also contain Y.* (Deenadayalan et al. 2014)

$$conf\left(x \Rightarrow y\right) = \frac{supp(XY)}{supp(X)}$$

Because in this project, the rules are deduced from patterns instead of transactions, the confidence of a rule denotes how often the action Y appears in pattern that contain the events X (condition). If a rule has a confidence of 100%, no occurrence of the pattern without the action was mined. The confidence is calculated using the support (count) of the pattern (Deenadayalan et al. 2014). The support count of the pattern including the action is divided through the support count of the pattern without the action. This leads to the formula in the following Figure 28.

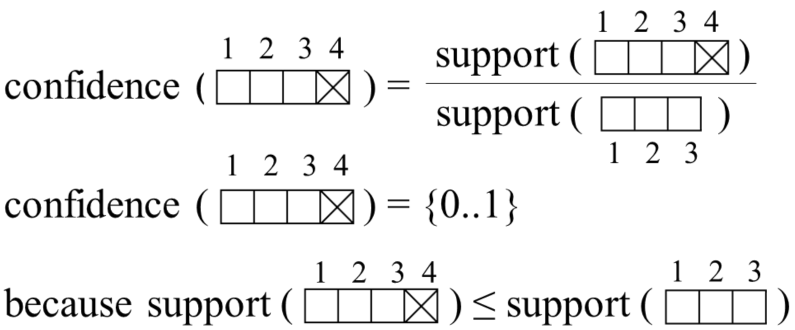


Figure 28 Formula to calculate the confidence of a rule

When mining overlapping patterns, the support count of a shorter pattern containing the same events as a longer pattern must be equal or higher than the longer, because each occurrence of the longer pattern will also increase the count of the shorter pattern. Therefore, the confidence calculated by this formula (Figure 28) will always result in the desired interval between zero and one. However, this approach is not working when the action is not at the beginning or at the end of a pattern. In this case, the displacement of the action creates a completely new pattern with a support count that is completely independent from the original one (Figure 29).

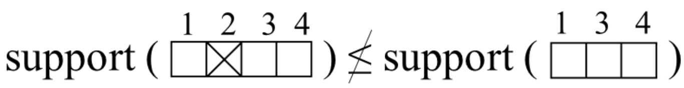


Figure 29 If the action is not the first or the last event, the support count is independent

Therefore, the formula introduced in Figure 28 cannot be used to calculate also the confidence for patterns where the action is not at the beginning or at the end of the pattern. The confidence would not result in the interval between zero and one and would therefore not be comparable to the other patterns. Even a division by zero can arise if the shorter pattern does not occur in the underlying data at all. To make the same statement as the formula in Figure 28 the formula has to be modified (Figure 30).

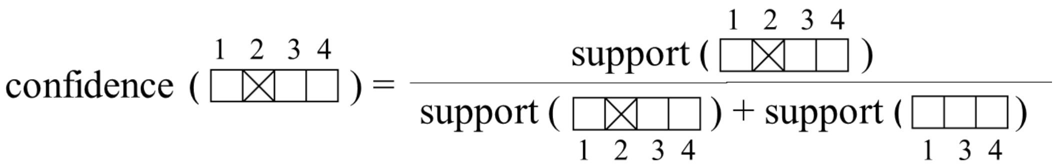


Figure 30 Adapted formula for confidence

To satisfy the need for one formula, which can be applied for both cases, the general formula was modified as follows.

$$conf\left(x \Rightarrow y\right) = \frac{supp(XY)}{supp(XY \cup Y)}$$

Due to this modification, no division through zero is possible and the confidence for both cases results in an interval between one and zero and therefore gets comparable. The support count of the divisor is increased each time X and Y (original pattern) or just X occurs. This results in the correct support count for both cases, because for patterns where the action is at the beginning or at the end $X \subset (X \cup Y)$ whereas for the other patterns $X \not\subset (X \cup Y)$.

The distribution of the confidence calculated for all relevant patterns mined in this project is illustrated in the histogram in Figure 31.

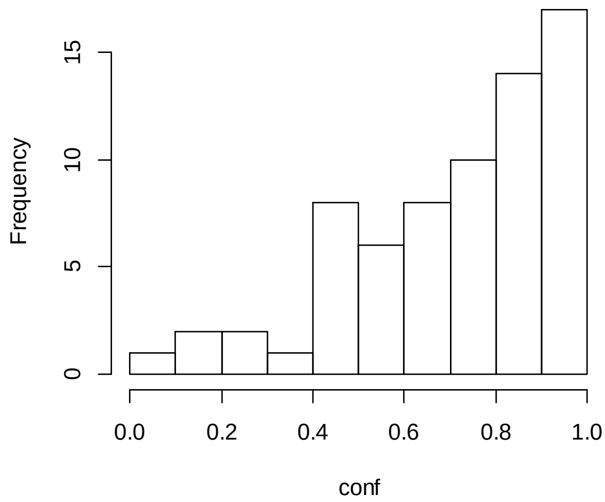


Figure 31 The histogram illustrates the distribution of confidence

7.4 SPECIFIC IMPLEMENTATION

This subchapter provides a closer insight in the specific implementation of the recommender system for this master thesis. To make an evaluation of the system in real smart homes possible, a fully functional prototype was implemented.

7.4.1 System

The recommender system is implemented on a cloud based Microsoft Azure VM (Virtual Machine) using IaaS (Infrastructure as a Service). The VM was set up with Ubuntu V14.04 and can be reached under the following URL http://dslivestat.cloudapp.net.

To run the recommender system, the following packets and python modules must be installed:

Packages

- python, python-dev
- mysql
- php5, php-mysql
- libxml2-dev, libxslt1-dev
- apache2, libapache2-mod-wsgi

Python modules

- matplotlib
- SQLAlchemy
- lxml
- Flask

7.4.2 Architecture

The smart homes event data is parsed from the log files of the DigitalSTROM system (example file in appendix: Smart home log file extract). The access to the log files was provided by Aizo. The files are uploaded by a script installed on the DigitalSTROM infrastructure in the houses and made accessible for this project on a file server (aizoupload.cloudapp.net). The files are copied all 5 minutes by Rsync (remote synchronization) to the VM where the recommender system is running. They are parsed by a python script and stored in a MySQL database (illustrated in Figure 32). In contrast to the general architecture (described in chap. 7.2), the system is not using one database per house but one central storage. Therefore, the recommender system must be able to distinguish the events and the patterns from the different houses.

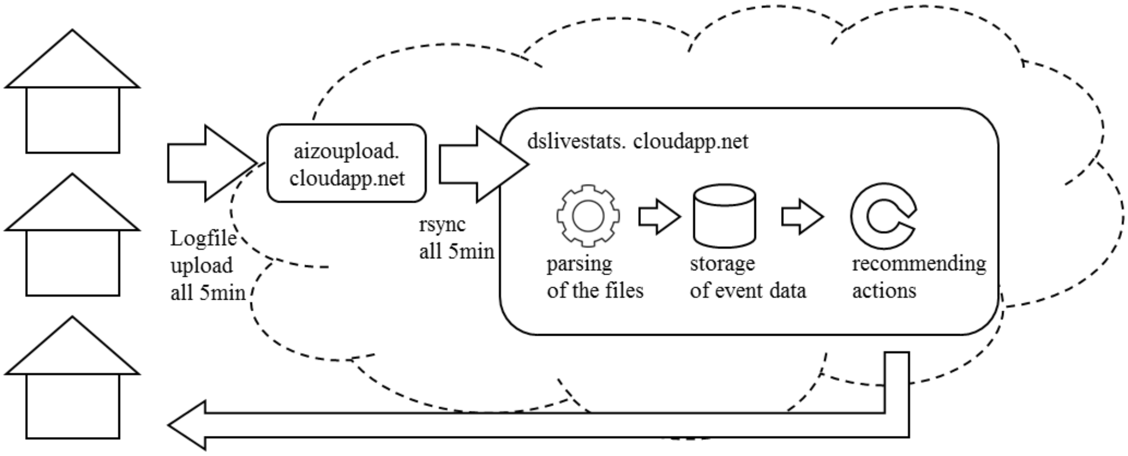


Figure 32 Specific implementation of the system using AIZO data

Because the timestamp for event data is only accurate on seconds, events happening in the same second have the same timestamp. To prevent that events disappear, they cannot be grouped just by timestamp but also by id. Furthermore, it is important to order the time series of events not only by timestamp but also by id. This sorting will not correct the missing accuracy of the timestamp, but it makes the order at least consistent, which is essential for the system.

The association rules are generated from the patterns during runtime, therefore the database is called pattern database. It is designed to store the relevant patterns mined in the underlying events. One installation can have zero to n patterns, which can consist of one to n events. The information about the position of the event in the pattern is stored in the "position" attribute (in the event table). Because it is important to know which events are actions, the attribute "is_action" was introduced on event level.

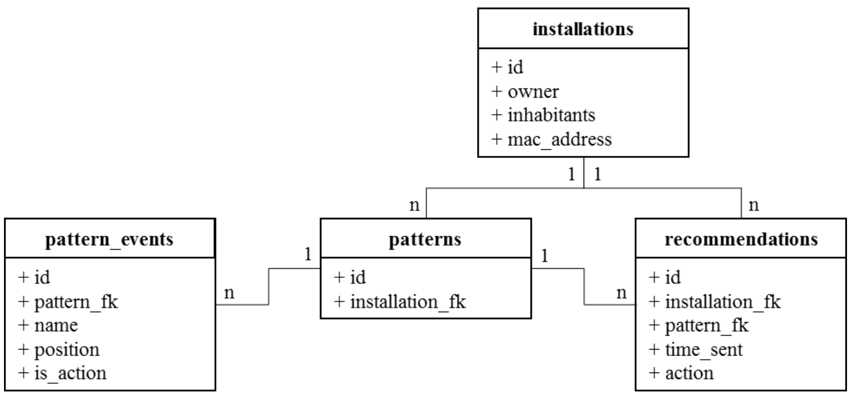


Figure 33 ERD of the pattern database of the recommender system

Another difference between the theoretical concept and the implementation of the prototype is that the events are not delivered in a real event stream but all five-minutes. Therefore, the events are parsed directly into the event database and the recommender is consuming the “event stream” from the database.

7.5 SUMMARY OF CHAPTER 6

In this chapter, the research question (RQ4) about the design of such a recommender system was answered, by the presentation of the concept and the implementation of the system as a fully functional prototype using the data of the digitalSTROM system.

The recommender system follows an association rules based approach to recommend actions to lower energy inside smart homes. The matching algorithm, which is used to apply the rules on the event stream, is based on a deterministic finite state machine. Furthermore, the calculation of confidence for the patterns mined in this project was introduced. The final formula developed for confidence of a rule is defined as follows:

$$conf\left(x \Rightarrow y\right) = \frac{supp(XY)}{supp(XY \cup Y)}$$

While the design of the recommender system, it became clear that there are two types rules, which must been treated differently. Not just for the implementation of the state machine but also for calculation of the confidence, one of the major challenges was to find an approach, which was working for both, patterns where the action was at the beginning or the end and patterns where the action is in the middle.

CHAPTER 8 - EVALUATION

8.1 INTRODUCTION OF CHAPTER 8

While the chapters 4, 5 and 6 deal with manual data analysis and the design and development of the recommender system, this seventh chapter is going to present the setup and the results of the evaluation of the recommender system. Furthermore, the aim of the chapter is to answer research question 5:

- How can be evaluated, whether a suggested action does not decrease comfort levels

To test this hypothesis of this master thesis, it is essential to realize a field test with real inhabitants. Therefore, a functional prototype of the system developed was implemented, which can be used for a field test. Since the recommender suggests only actions that lower energy usage, the first part of the hypothesis is not difficult to prove. However, to prove if an action suggested does decrease comfort levels is much more difficult. Therefore, the inhabitants of the test smart homes must decide, if an action suggested for their home did decrease comfort levels or not.

8.2 SETUP

The evaluation is conducted in real households with real inhabitants producing real event data. The test households were provided by a Swiss smart home system provider called Aizo. Therefore all test houses were equipped with the smart home automation system DigitalSTROM developed by Aizo. Overall, 33 households were considered for the evaluation. The historical event data of the installations (underlying data) was mined by the recommender system in advance. According to the number of relevant patterns found in the houses, they were ranked and the owners of the 15 most promising households where requested to take part in the evaluation. Finally, the owners of eight houses agreed to participate in the evaluation of this project. Fortunately both, single- and multi-inhabitant houses where represented.

The setup of the evaluation was defined, so that the interval between the events causing the recommender to send a recommendation and the actual reception of the recommendation is as short as possible, because the action suggested is only valid for a certain time. Furthermore, the aim was to make the assessment of a suggestion as simple as possible, to encourage the inhabitants to vote and to reach a high degree of answered suggestions. On that account, the recommendations were sent per Short Message Service (SMS) to the mobile devices of the inhabitants. The evaluation process was designed, that multiple contacts could be provided per installation (illustrated in the ERD in Figure 34). A recommendation was sent to all contacts referenced to the affected installation. To ensure, that no data will be lost, the feedback from all contacts was saved. Thus, it can be decided during evaluation how to interpret the given results.

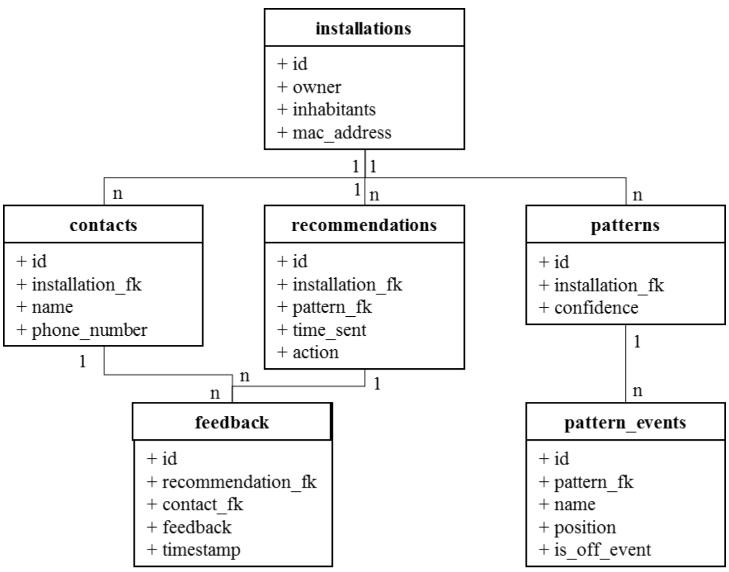


Figure 34 Relations of the entities involved in the evaluation process

The inhabitants were briefed before the evaluation about the purpose and the proceeding of the evaluation. The guidance distributed to the participants can be found in the Appendix (Guidance for evaluation participants). To judge if a recommendation does not decrease comfort levels, inhabitants where asked to decide if a recommendation is “useful”, while useful was defined in the guidance for the evaluation as follows:

A useful recommendation is defined as a recommendation "which will lower energy usage but not decrease comfort levels".

Besides the recommended action and a timestamp, the message sent to the inhabitants contains two hyperlinks, which enables a comfortable and simple voting. An example of such a text message can be seen in Figure 35.

```
Hi Michael Zehnder,

I would recommend to turn Off-Device
Nachttischlampe (on 2014-11-16 23:19:16).

Is this recommendation useful?

Yes: http://snurl.com/29fgj3d?r=2663&c=4f=1
No:  http://snurl.com/29fgj3d?r=2663&c=4f=0
```

Figure 35 Example of a text message sent to the inhabitants

8.3 PHASE 1

Because in design research, the hypothesis is rarely proven by the first prototype, the evaluation is structured in two independent phases while two separate periods. The aim of the first phase is to provide a large basis of data for evaluation and further improvement of the system. The analysis of the data collected during phase 1 should help to improve the recommender system in terms of decreasing the negatively rated recommendations in phase 2, while holding the positives at the highest amount possible. Furthermore, the first phase should prove the concept of the recommender system, which implies one useful recommendation as an absolute minimum. To reach this goal, a setup was chosen where many recommendations were sent, while having the awareness that this setup could also lead to many negative feedbacks.

Furthermore, one restriction was introduced to protect the inhabitants from too many wrong recommendations: If a rule causes 10 negative feedback in a row, no more recommendations caused by this rule will be sent.

8.3.1 Results phase 1

The first test phase started at December 5th 2014 and ended at December 19th 2014. Because of the fact, that the test period lasted exactly two weeks, all weekdays occur in equal amount. The key results of the first phase are summarized in the following Table 11.

Number of days evaluated	14
Recommendations sent	160
Number answered recommendations	76
Number of recommendations voted useful	7
Number of recommendations voted not useful	69
Ratio answered recommendations / recommendations sent	47.5%
Ratio useful recommendations / answered recommendations	9.21%
Number of active rules	54
Number of rules caused recommendations	23
Number of rules reached 10 negative feedbacks in a row	5

Table 11 Key results of the evaluation in the first phase

8.3.2 Findings phase 1

The findings gained from the results of the first phase of the evaluation can be divided in two parts. One part of the findings came from the evaluation data itself and a second part was deducted from interviews with the inhabitants.

User feedback findings
The feedback of the interviews with the inhabitants revealed two main findings:

- Absent scene should not be recommended as action
- The low response rate of 47.5% was caused by ambiguous recommendations

The "absent scene" is a predefined scene of the DigitalSTROM system, which allows turning off all lights when leaving the house. This is normally realized by a toggle button next to the entrance, which can trigger "Absent" and "Present" scenes alternately. As described in state machine in Figure 27, a recommendation is only being sent, if the condition of the rule matches and the subsequent event is not the action itself. Therefore, after the match of a rule, one event must be awaited to exclude the possibility that it will be the action recommended. Because of this fact, the recommendation for an "absent scene" makes no sense even if the inhabitant did forget to activate the absent scene before the leaving the house, because it will be sent not until the inhabitant is back inside the house and causes another event.

When questioned to the low response rate, some test persons mentioned, that they were not able to answer recommendations because they were ambiguous. As described in chapter 4.4.1, inhabitants can name scenes and devices individually and therefore duplicates are possible. During the interviews, it became apparent, that some users gave similar or equal names for different devices inside their household. Consequently, a recommendation to turn off the "spot light" could apply for more than one device and they could not decide if the recommendation was useful.

Evaluation data findings
The goal of the analysis of the evaluation data was to find characteristics of rules that give indications about its feedback. Such attributes can already be considered during the filtering process of the rules (in phase 2) and lead to better rules and consequently to better recommendations.

For the analysis, all feedbacks were grouped by rule. To get an indication of the "usefulness" of a rule, the feedback was weighted as follows:

Feedback	Weight
Useful (positive)	+1
Not useful (negative)	-1
No feedback	0

Table 12 Weight of the feedback

All rules that did not cause any recommendations were not considered, because nothing about the feedback of these rules can be said. All 31 rules that caused no recommendations would have resulted with a weight of zero, what would have distorted the result of the analysis.

The distribution of the weighted feedback over the 23 rules that caused recommendations is presented in form of a histogram in Figure 36. All rules result in weighted feedback between -12 and +2, which is showed on the x-axis. The frequency (y-axis) displays the number of rules, which resulted in the given interval. Because the weighted feedback is not perfectly normal distributed, but rather skewed to the left, the results of the following regression analysis have to be handled with particular caution. Because transforming the distribution into an approximation of the normal distribution (Schmidt 2009) is beyond the scope of this paper, the regression analysis is proceeded with the data at hand.

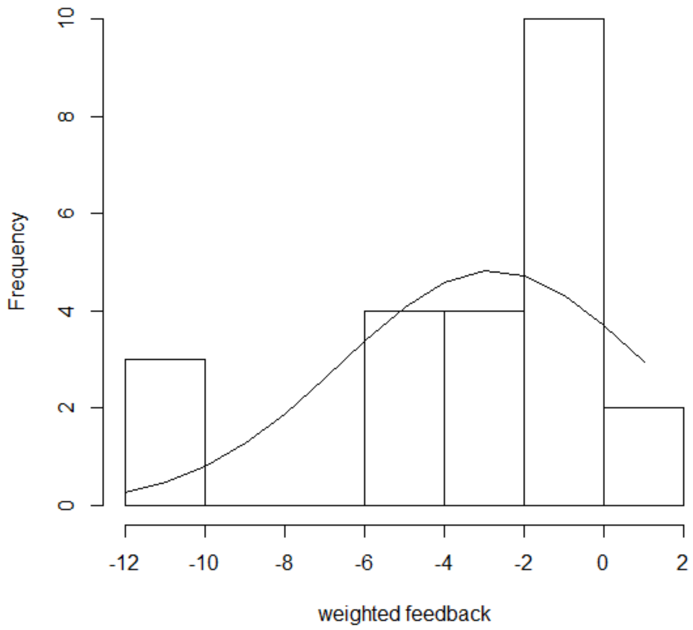


Figure 36 Histogram of feedback with a normal curve

For the following regression analysis, the dependent variable is the “weighted feedback” and the attributes introduced in chapter 7.3 for the prioritization served as explanatory variables:

- The length of the pattern
- The position of the action
- Support of the pattern
- Confidence of the rule

Because repeating the pattern mining on the event data barely produced new patterns, the date when the rule was mined was ignored as explanatory variable. All other attributes were analysed individually as simple linear regression but also as multiple regression. The according scatter plots (incl. regression line) are presented in the following figures Figure 37 and Figure 38.

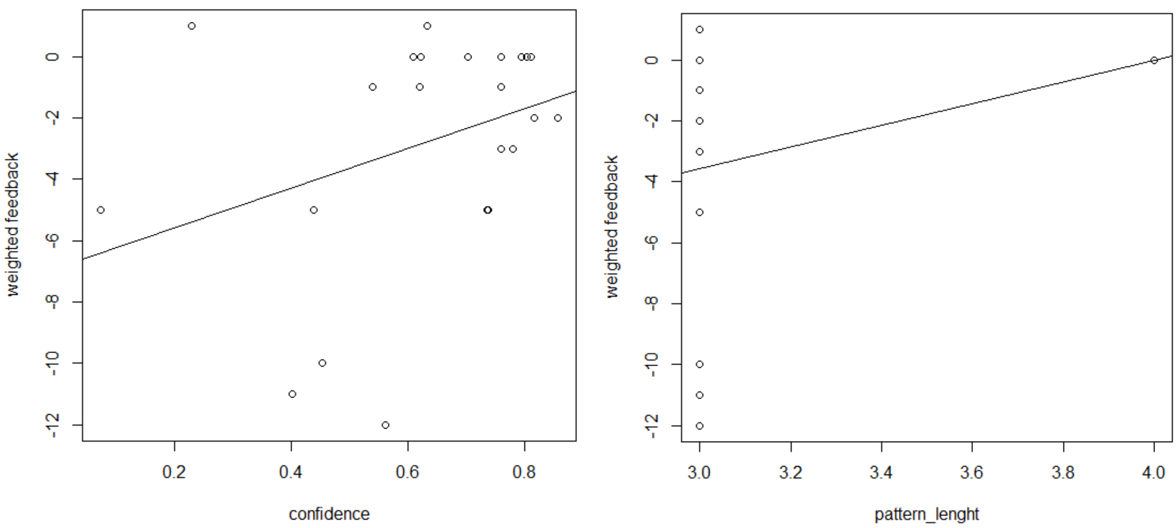


Figure 37 Scatter plot feedback / confidence and feedback / pattern_ length with line of best fit

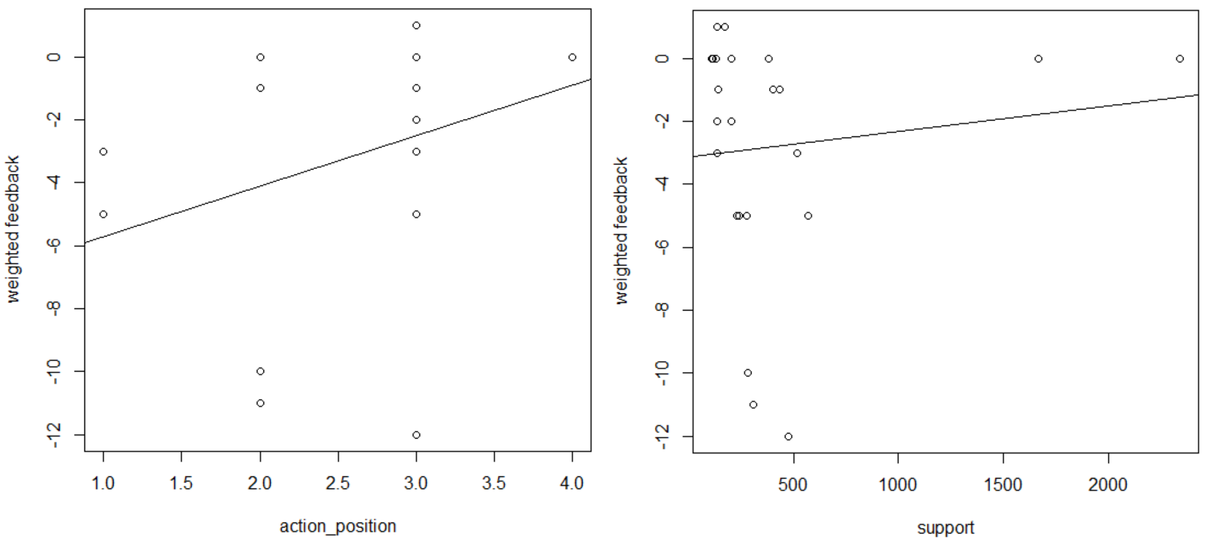


Figure 38 Scatter plot feedback / action_position and feedback / support with line of best fit

A more detailed summary of the simple linear regression models is given for each explanatory variable in Table 13. A detailed explanation of the methods and a definition of the terms used, can be found in “Linear Regression and Correlation” (Kenney & Keeping 1962).

variable	**estimate**	**std. error**	**t-value**	**p-value**
confidence	6.503	3.906	1.665	0.04941
pattern_length	3.556	1.807	1.967	0.0625
action_position	1.6106	0.8668	1.858	0.07724
support	0.000801	0.001560	0.514	0.61291

Table 13 Detailed summary of the simple linear regression analysis

The summary reveals, that confidence has a p-value smaller than 0.05 and therefore is significant for the dependant variable feedback. In contrast, the regression model using the variable “support” shows no significance. Furthermore, all possible combinations of the variables were analysed using multiple regression models. The regression model including confidence and pattern_length as explanatory variables scored the lowest p-value (Figure 39).

```
Residuals:
    Min      1Q  Median      3Q     Max
-9.1380 -1.3394  0.7033  2.1423  5.2264

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       -8.7632     2.9695  -2.951   0.0079 **
confidence         4.1310     4.3130   0.958   0.3496
action_position    1.1954     0.9707   1.231   0.2324
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.613 on 20 degrees of freedom
Multiple R-squared:  0.1789,    Adjusted R-squared:  0.09674
F-statistic: 2.178 on 2 and 20 DF,  p-value:0.04418
```

Figure 39 Summary of the multiple regression model by R

Because of the results of this regression analysis, the following hypothesis was formulated: Patterns with high confidence and high pattern-length tend to receive better feedback then the other patterns. On the other hand, support (count), which was the major attribute for mining frequent patterns, did not show any significance to describe the feedback a rule. To validate the hypothesis formulated with figures, the results from phase 1, analysed a second time with the focus on pattern length and confidence.

8.3.3 Confidence and pattern length analysis

To receive a figure for the "usefulness" of a rule, confidence and pattern-length of each rule where multiplied with their estimate (Figure 39) and add together to calculate a single coefficient for each rule. The rules resulted in coefficients between 1.2 and 5.33. Consequently, an additional step in the relevant pattern filter was created, which does reject patterns resulting in a coefficient under a given threshold.

To prove the result of the regression analysis and to find a suitable threshold for the second evaluation phase, the data from phase 1 was taken, and the "useful recommendations to answered recommendations ratio" (ratio) was simulated for each possible threshold of the "confidence and pattern-length coefficient" (coefficient) between 1.2 und 5.4 (accurate to a tenth). The result is visualized in Figure 40.

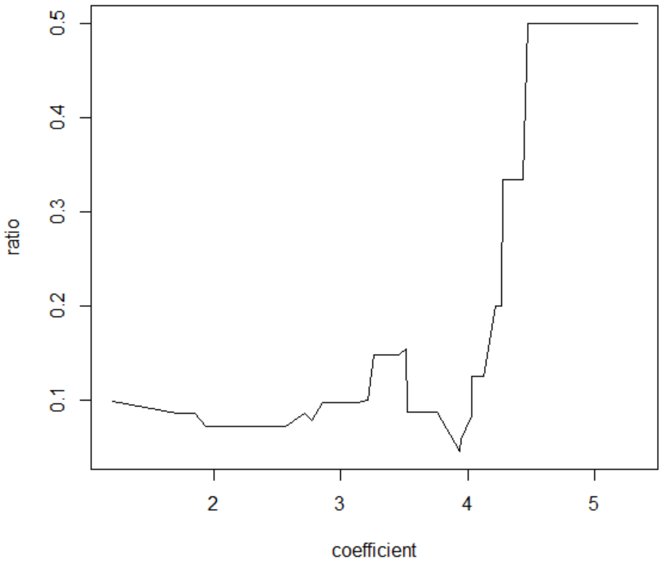


Figure 40 Line chart of useful recommendations to answered recommendations ratio per coefficient

In Figure 40, it becomes obvious, that the coefficient might be a suitable indicator for feedback. The ratio of useful recommendations compared to answered recommendations (ratio), rises together with the coefficient. However, the dependency between the two variables is not linear, after 3.7 the ratio drops even under the initial value of 9.21% (which was the result achieved in phase 1 without any restrictions by a coefficient). Because of this local maximum at a coefficient value of 3.7, it became the threshold for the second phase. It should allow achieving better results, without excluding too many rules.

8.4 PHASE 2

The aim of the second phase of the evaluation was to increase the ratio of answered recommendations and the ratio of useful recommendations compared to the first phase. The analysis of the data collected during phase 1 was used to adapt the system, which should lower negative rated and unanswered recommendations. The prototype was improved in the following points:

- At first, all rules which were excluded during phase 1, because of 10 negative feedbacks in a row, were not considered in phase two anymore. To make sure that the second phase remains independent, the feedback count from phase 1 was reset for all patterns.
- Furthermore, all rules which recommended the action "absent", where excluded from the second phase because of the user feedback after phase 1 (chap. User feedback findings).
- To reduce the problem of the low response rate caused by ambiguous recommendations, the recommender system was upgraded between phase 1 and 2. Because the information about the location of the devices was not part of the underlying data, it was directly parsed from the

apartment.xml from the DigitalSTROM system. Due to this feature, the recommendations were enriched by the room of the device, which is apparent in the following example (Figure 41). Since no manual matching was necessary, this feature is still in the scope of this project.

- As result of the regression analysis, confidence and pattern-length of each rule where multiplied with their estimate (Figure 39) to calculate a coefficient which gives indication about the "usefulness" of a rule. A threshold of 3.7 was defined and all rules with a coefficient below this value where excluded from the second phase (explained in chap 8.3.3).

```
Hi Michael Zehnder,

I would recommend to turn Off Raumtaster in
Waschküche (on 2015-01-08 17:49:18).

Is this recommendation useful?

Yes: http://snurl.com/29fgj3d?r=2663&c=4f=1
No:  http://snurl.com/29fgj3d?r=2663&c=4f=0
```

Figure 41 Example of a text message sent to the inhabitants in phase 2

8.4.1 Results phase 2

The second test phase started at January 5th 2014 and ended after 14 days at January 19th 2014. Because of the fact, that the test period lasted exactly two weeks, all weekdays occur in equal amount. The key results of the first phase are summarized in Table 14.

Number of days evaluated	14
Recommendations sent	18
Number answered recommendations	9
Number of recommendations voted useful	1
Number of recommendations voted not useful	8
Ratio answered recommendations / recommendations sent	50%
Ratio useful recommendations / answered recommendations	11,11%
Number of active rules (patterns)	35
Number of rules (patterns) caused recommendations	6
Number of rules reached 10 negative feedbacks in a row	0

Table 14 Key results of the evaluation phase two

8.4.2 Findings phase 2

Although the differences between phase 1 and phase 2 are rather smaller than expected, both aims, which were to increase the ration of answered recommendations and the ratio of useful recommendations where completed successfully. The ratio of useful recommendations was increased by 1.9% (from 9.21% in the first phase to 11.11% in the second). Furthermore, the ratio of answered recommendation could be increased by 2.5% (from 47.5 to 50%), which is still lower than expected. Further questions to the inhabitants could not determine why the answered recommendations where still low in the second phase. It is assumed that the discipline to answer every recommendation might have deteriorated after four weeks of testing.

Because of the low response rate and the reduced number of rules, the two weeks provided too little results for significant findings. To achieve results, which are more significant, the evaluation phase must have been extended, which would have outran the scope of this project.

However, the results of the second phase confirm the assumptions made after phase 1. These findings are summarized in the following chapter 9.

8.5 SUMMARY OF CHAPTER 8

In this eighth chapter, the question "how can be evaluated, whether a suggested action does not decrease comfort levels" is answered. A setup for the evaluation was developed and the functional prototype of the recommender system designed and implemented in this project was installed in eight single- and multi-inhabitant smart homes occupied by volunteer testers from Aizo. To measure if a suggestion does decrease the comfort levels or not, the inhabitants where advised to vote for each recommendation.

The results of the first phase of the evaluation proved that the concept of the recommender system does work in reality. The system produced seven useful recommendations to save energy during the first 14 days of testing in the eight test houses. The percentage of negative recommendations, considered as not useful, has been decreased slightly in the second phase (1.9%).

Furthermore, it was exposed that confidence and pattern length can give an indication about the feedback of a rule. They can be used not only to prioritise a rule, but also to exclude a rule from being executed by the recommender.

CHAPTER 9 - CONCLUSION

9.1 INTRODUCTION OF CHAPTER 8

In this final chapter, a summary of the project is made and the main findings of the project are presented. In this context, the last research question (RQ 6) is answered:

- Do the suggested actions decrease comfort levels?

In addition, the thesis statement of this project is discussed and proven. Finally, an outlook for possible future research, based on the findings of this project, is given.

9.2 SUMMARY OF FINDINGS AND CONTRIBUTION

The following subchapters contain the main findings contributed by this project. Because the scope of this project was set in order to cover as many smart home scenarios as possible (no motion detection devices, no user interaction, multiple resident capability), most of the findings can be applied for smart homes in general, even if the system was only tested using the data from the smart home automation system DigitalSTROM by Aizo.

9.2.1 Relevant actions to lower energy usage

This project succeeded in developing a prototype of a recommender system to suggest relevant actions for energy saving in smart homes based on inhabitants' behaviour data only. Therefore six research questions/objectives where addressed and answered. The following paragraphs give a summary of the findings of the project structured by research objectives, which are deducted from the research questions.

To identify possible actions, which can be suggested in smart homes to lower energy usage.
The project succeeded in identifying possible actions, which can be suggested to lower energy usage in smart homes. It was de exposed, that these actions must be a subset from all events occurring in the smart home, to be suggested by an autonomously and self-adapting recommender system. Actions differ from other events because they are energy saving. Criteria must be defined to segment the actions from the (normal) events according to their attributes. The criteria might differ, depending on the underlying data. For this concrete case, the events where identified as actions, when they matched one of the following criteria.

- The name of the scene called by the event does contain one of these key words: Dim, Off, Sleep, Standby Stop.
- The name of the type of the sensor event, which was reported by the event, does contain one of these key words: Aus, Off, Uit.

To find algorithms which can be used to find relevant patterns in the behaviour data.
It was explored, that patterns in the behaviour data of the inhabitants can be used to trigger the actions identified at the right moment in the smart homes. Therefore, different algorithms for frequent (and/or periodic) pattern mining in sequential smart home behaviour data where evaluated as basis for the relevant pattern mining algorithm. Based on the findings of Schweizer (2014), the algorithm WSDD was used to mine frequent and periodic patterns for the prototype of the recommender system.

To mine only relevant patterns, WSDD was enhanced by a relevant pattern filter. The filter extracts the patterns relevant for suggesting actions from the results set of WSDD.

To define which types of behaviour patterns are needed to suggest actions.
To make a distinction between the relevant patterns and (normal) frequent and periodic patterns, the output of WSDD (or any other pattern mining algorithm used) must be post filtered. To enable the recommender system to filter the relevant patterns without user interaction, filter criteria must be defined. To be considered as relevant, a pattern must fulfil all of the following criteria:

- Pattern must contain at minimum one action to lower energy usage.

- Minimum pattern length of three events.
- For frequent patterns: minimum support of 0.05.
- For periodic patterns: minimum 75% of the timespans between the occurrences of the pattern are isochronous (with a grace period of 10%, defined in chap 5.2).
- Minimum confidence and pattern length coefficient of 3.7 (defined in chap. 9.2.2)

To design a recommender system.
Furthermore, the design for a recommender system using these patterns to suggest actions was proposed and implemented. Therefore, the patterns where translated into association rules and matched against the real time event feed of the smart homes. An association rule to lower energy in smart homes consists of two parts: The relevant pattern without the action (condition) and the action to lower energy (action):

- {condition event 1, condition event 2, condition event n} ⇒ {action}

The matching of the association rules is designed as state machine. To prioritize the matching of the rules, confidence is used (chap. 9.2.2).

To evaluate if the suggested actions do decrease comfort levels
To evaluate if an action suggested is useful for the inhabitant and did not decrease comfort levels, the setup for a field test was developed. Because the company Aizo offered to provide test houses, a prototype of the recommender was developed using the data from the DigitalSTROM smart home automation system developed by Aizo. The prototype was installed in eight test households and tested for an overall period of 1 month. The recommendations made by the system where sent to the inhabitants' mobile devices and rated by them. During this test period, nine successful suggestions were made by the system, which were rated by the inhabitants as energy saving but not decreasing comfort levels.

Proof of the thesis statement
For repetition, the thesis statement of this thesis is given as follows:

It is possible to design a recommender system that can suggest actions in smart homes based on consumer behaviour, which will lower energy usage but not decrease comfort levels.

During the evaluation, all restrictions discharged from the thesis statement where observed:

- No additional sensors, motion and occupancy detection or tagged gloves and RFID-tags were installed to determine the location of the inhabitants (since this could decrease comfort levels).
- Not configuration or other interaction between the system and the inhabitant was necessary before or while the runtime of the recommender system (since this could decrease comfort levels).
- The only input data of the recommender system was the consumer behaviour data provided by the smart home automation system.

Because all these restrictions where fulfilled and the system managed to suggest actions, which did lower energy usage but not comfort levels, it could verify part of the statement.

However, the recommender system is still far from suggesting solely actions, which will lower energy usage but not decrease comfort levels. The ratio of useful recommendations, which reached little over 11% during the final test of the prototype, must be increased before broader adaption of the system is possible. Nevertheless, the proof of concept provided by the prototype was the first important step for further research in this field. Based on the findings of this master thesis, other research topics can be explored (chap. 9.3).

9.2.2 Confidence and pattern length more significant as support

The results of the evaluation of the prototype were analysed using simple and multiple regression models. The feedback of the inhabitants (if the recommendation was energy saving and did not lower energy usage) was the depending variable while the following attributes served as explanatory variables.

- The length of the pattern
- The position of the action
- Support of the pattern
- Confidence of the rule

It was discovered, that the confidence of the association rule discharged from the pattern is considerably more significant than the support of a rule. Furthermore it was discovered, that the most significant combination of attributes for the feedback of the suggestions provided by a rule is confidence and pattern-length, which resulted in a p-value of 0.04941.

Because of this findings, it is supposed, that the frequency and periodicity of a pattern are little important for the "usefulness" of a pattern. Therefore a recommender system using an algorithm, which mines pattern with high confidence, independent if they are frequent or periodic, might achieve better results. This idea is perused in the following sub chapter (9.3).

9.3 OUTLOOK AND NEXT STEPS

In the same manner as this master thesis builds on the findings of a master thesis called "Learning frequent and periodic usage patterns in smart homes" (Schweizer 2014), follow-up research projects can build upon the findings of this master thesis. To be more specific, different ideas for further research are presented. These ideas occurred during the design, implementation or evaluation of the recommender system and would have gone beyond the scope or the time restrictions set for this project.

- Using confidence and pattern length instead of support or periodicity as criteria for the mining algorithm to provide patterns in sequential smart home event data might succeed in more and in better patterns.
- The time between two events (or the action) is not considered neither by the mining algorithm nor by the recommender system. Using this information might improve the accuracy of the suggestions made by the system.
- Besides confidence, pattern-length and action position, other attributes could be introduced to decide if a rule is concerned or not. Such attributes might be:
    - Time of day when the pattern occurs most
    - Weekday when the pattern occurs most
    - Season when the pattern occurs most

  Such attributes could help to react on different behaviour patterns in different times and allow better conclusions than just the date when the rule was mined (e.g. a pattern, which occurred often exactly 1 year ago on the same weekday as today could lead to better results as a pattern mined recently). This could be achieved by an additional clustering step.
- The recommender could learn from the feedback of the inhabitants (and not just exclude the rule after 10 negative feedbacks in a row).

BIBLIOGRAPHY

Achar, A., Laxman, S. & Sastry, P.S., 2011. A unified view of the apriori-based algorithms for frequent episode discovery. *Knowledge and Information Systems*, 31(2), pp.223–250. Available at: http://link.springer.com/10.1007/s10115-011-0408-2 [Accessed May 28, 2014].

Alam, M.R. et al., 2012. SPEED: An Inhabitant Activity Prediction Algorithm for Smart Homes. , 42(4), pp.985–990.

Augusto, J. & Nugent, C., 2006. *Designing smart homes: the role of artificial intelligence*, Available at: http://books.google.com/books?hl=en&lr=&id=QHGHoZ53OBIC&oi=fnd&pg=PA1&dq=Designing+Smart+Homes+-+The+role+of+artificial+Intelligence&ots=bo2g68IRiS&sig=iyg6wwYaDDtdjvfxuv2cnE3cwfo [Accessed April 17, 2014].

Aztiria, A. et al., 2011. Discovering frequent user--environment interactions in intelligent environments. *Personal and Ubiquitous Computing*, 16(1), pp.91–103. Available at: http://link.springer.com/10.1007/s00779-011-0471-4 [Accessed June 4, 2014].

Björkskog, C., 2007. Human Computer Interaction in Smart Homes. *Helsinki, Finland*. Available at: http://www.hiit.fi/~oulasvir/58307110/smarthomes.pdf [Accessed March 20, 2014].

Blumendorf, M., 2013. Building sustainable smart homes. *on Information and Communication …*, pp.151–158. Available at: http://e-collection.library.ethz.ch/eserv/eth:6558/eth-6558-01.pdf#page=157 [Accessed April 13, 2014].

Boton-Fern´ndez, V. & Lozano-Tello, A., 2011. Learning Algorithm for Human Activity Detection in Smart Environments. *2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, pp.45–48. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6040701 [Accessed June 1, 2014].

Brush, A., Lee, B. & Mahajan, R., 2011. Home automation in the wild: challenges and opportunities. *Proceedings of the …*. Available at: http://dl.acm.org/citation.cfm?id=1979249 [Accessed March 29, 2014].

Chen, L., Nugent, C.D. & Wang, H., 2012. A Knowledge-Driven Approach to Activity Recognition in Smart Homes. *IEEE Transactions on Knowledge and Data Engineering*, 24(6), pp.961–974. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5710936.

Clinckx, N., 2013. Smart Home: Hope or hype? , (January), pp.1–20.

Crandall, a. S. & Cook, D., 2008. Attributing events to individuals in multi-inhabitant environments. *4th International Conference on Intelligent Environments (IE 08)*, pp.4B4–4B4. Available at: http://digital-library.theiet.org/content/conferences/10.1049/cp_20081164.

Das, S.K. et al., 2002. The role of prediction algorithms in the MavHome smart home architecture. *IEEE Wireless Communications*, 9, pp.77–84.

Deenadayalan, E., Hanirex, D. & Kaliyamurthie, K., 2014. ASSOCIATION RULE MINING IN DISTRIBUTED DATABASE SYSTEM. , 3(4), pp.727–732. Available at: http://ijcsmc.com/docs/papers/April2014/V3I4201499a68.pdf [Accessed January 24, 2015].

Douven, I., 2011. Abduction. *Stanford Encyclopedia of Philosophy*. Available at: http://plato.stanford.edu/entries/abduction/.

Du, Y. et al., 2006. Recognizing Interaction Activities using Dynamic Bayesian Network. *18th International Conference on Pattern Recognition (ICPR'06)*, pp.618–621. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1698968.

Energie Schweiz für Gemeinden, Stadt Zürich & SIA Schweizerischer Ingenieur- und Architekten verein, 2012. Bilanzierungskonzept 2000-Watt-Gesellschaft.

Esmaeili, M. & Gabor, F., 2010. Finding sequential patterns from large sequence data. *arXiv preprint arXiv:1002.1150*, 7(1), pp.43–46. Available at: http://arxiv.org/abs/1002.1150 [Accessed January 24, 2015].

Federal Administration, 2012. Families, households – Data, indicators. Available at: http://www.bfs.admin.ch/bfs/portal/en/index/themen/01/04/blank/key/haushaltsgroesse.html.

Forgy, C.L., 1982. Rete : A Fast Algorithm for the Many PatternIMany Object Pattern Match Problem *. , 19(1982), pp.17–37.

Gopalratnam, K. & Cook, D., 2007. Online Sequential Prediction via Incremental Parsing: The Active LeZi Algorithm. *IEEE Intelligent Systems*, 22(1), pp.52–58. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4078956.

Harper, R., 2003. *Inside the Smart Home* R. Harper, ed., London: Springer-Verlag. Available at: http://link.springer.com/10.1007/b97527.

He, Y., 2010. Energy saving of central air-conditioning and control system: Case study: Nanchang Hongkelong Supermarket. Available at: http://theseus17-kk.lib.helsinki.fi/handle/10024/21077 [Accessed May 4, 2014].

Inji, E., Attia, I. & Hamdy, P., 2011. Energy Saving Through Smart Home. , (2), pp.223–227.

Jahn, M. et al., 2010. The Energy Aware Smart Home. , pp.0–7.

Jain, G., Cook, D. & Jakkula, V., 2006. Monitoring health by detecting drifts and outliers for a smart environment inhabitant. *... conference on smart homes and health ...*. Available at: http://www.eecs.wsu.edu/~cook/pubs/icost06.pdf [Accessed March 20, 2014].

Jakkula, V., Youngblood, G. & Cook, D., 2006. Identification of lifestyle behavior patterns with prediction of the happiness of an inhabitant in a smart home. *... Approaches to Beauty and Happiness*. Available at: http://www.aaai.org/Papers/Workshops/2006/WS-06-04/WS06-04-005.pdf [Accessed March 20, 2014].

Johnson, P. & Clark, M., 2006. *Business and management research methodologies*, Auerbach Publications.

Kenney, J.F. & Keeping, 1962. *Linear Regression and Correlation,*

King, N., 2003. Smart home–a definition. *Intertek Research & Testing Centre*, pp.1–6. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:SMART+HOME+–+A+DEFINITION#5 [Accessed May 11, 2014].

Korvemaker, B. & Greiner, R., 2000. Predicting Unix command lines: adjusting to user patterns. *AAAI/IAAI*. Available at: http://www.aaai.org/Papers/Symposia/Spring/2000/SS-00-01/SS00-01-013.pdf [Accessed May 25, 2014].

Krishnan, N., Cook, D. & Wemlinger, Z., 2013. Learning a taxonomy of predefined and discovered activity patterns. *Journal of Ambient Intelligence and …*. Available at: http://iospress.metapress.com/index/T7072485H05H4707.pdf [Accessed June 5, 2014].

LeMay, M., Haas, J. & Gunter, C., 2009. Collaborative recommender systems for building automation. *System Sciences, 2009. HICSS …*, pp.1–10. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4755525 [Accessed June 5, 2014].

Li, Z. et al., 2010. Mining periodic behaviors for moving objects. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '10*, p.1099. Available at: http://dl.acm.org/citation.cfm?doid=1835804.1835942.

Manilla, H., Toivonen, H. & Verkamo, A., 1995. Discovering frequent episodes in sequences. *... of the First International Conference on …*. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Discovering+frequent+episodes+in+sequences#1 [Accessed March 29, 2014].

Mannila, H., Toivonen, H. & Verkamo, A.I., 1997. Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery*, 1, pp.259–289. Available at: http://www.springerlink.com/index/R0340461M0V78158.pdf\nhttp://link.springer.com/article/10.1023/A:1009748302351.

MavHome, 2003. MavHome::Publications. Available at: http://ailab.wsu.edu/mavhome/publications.html.

Nazerfard, E., Rashidi, P. & Cook, D., 2011. Using association rule mining to discover temporal relations of daily activities. *Toward Useful Services for Elderly and …*, pp.49–56. Available at: http://link.springer.com/chapter/10.1007/978-3-642-21535-3_7 [Accessed June 7, 2014].

Nowak, M. & Urbaniak, A., 2011. Utilization of intelligent control algorithms for thermal comfort optimization and energy saving. *2011 12th IEEE Control Conference*, pp.270–274. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5945862.

Peng, W.-C. & Liao, Z.-X., 2009. Mining sequential patterns across multiple sequence databases. *Data & Knowledge Engineering*, 68(10), pp.1014–1033. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0169023X09000573 [Accessed January 24, 2015].

Rashidi, P. et al., 2011. Discovering Activities to Recognize and Track in a Smart Environment. *IEEE transactions on knowledge and data engineering*, 23(4), pp.527–539. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3100559&tool=pmcentrez&rendertype=abstract.

Rashidi, P. & Cook, D., 2013. COM: A method for mining and monitoring human activity patterns in home-based health monitoring systems. *ACM Transactions on Intelligent Systems and …*, 9(4). Available at: http://dl.acm.org/citation.cfm?id=2508045 [Accessed June 5, 2014].

Rashidi, P. & Cook, D., 2010. Mining sensor streams for discovering human activity patterns over time. *Data Mining (ICDM), 2010 IEEE 10th …*. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5693997 [Accessed June 26, 2014].

Rashidi, P. & Cook, D.J., 2010. Mining and Monitoring Patterns of Daily Routines for. , pp.336–345.

Rashidi, P., Member, S. & Cook, D.J., 2009. Keeping the Resident in the Loop : Adapting the Smart Home to the User. , 39(5), pp.949–959.

Ricci, F., Rokach, L., Shapira, B., et al., 2011. *Recommender Systems Handbook* F. Ricci et al., eds., Boston, MA: Springer US. Available at: http://link.springer.com/10.1007/978-0-387-85820-3 [Accessed July 9, 2014].

Ricci, F., Rokach, L. & Shapira, B., 2011. *Recommender Systems Handbook* F. Ricci et al., eds., Boston, MA: Springer US. Available at: http://link.springer.com/10.1007/978-0-387-85820-3 [Accessed May 24, 2014].

Rohini, P. & RajKumar, R., 2014. International Journal of Advanced Research in Computer Science and Software Engineering A New Approach to Behavioral Reasoning in Smart Homes using DVSM Algorithm. , 4(2), pp.334–341.

Ruvini, J.-D. & Dony, C., 2000. Learning Users' Habits to Automate Repetitive Tasks. In *Your Wish is My Command: Programming by Example*. pp. 271–296. Available at: http://web.media.mit.edu/~lieber/Your-Wish/14-Ruvini.pdf.

Saunders, M. et al., 2009. Research methods for business students, 5/e. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Research+Methods+for+Business+Students#0 [Accessed June 1, 2014].

Schmidt, A., 2009. Normalverteilungsannahme und Transformationen bei Regressionen. ... *Aspekte lose gekoppelter Systeme und Electronic ...*, pp.1–17. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Normalverteilungsannahme+und+Transformationen+bei+Regressionen#0 [Accessed January 7, 2015].

Schweizer, D., 2014. Learning frequent and periodic usage patterns in smart homes. , (January).

Sequeira, K. & Zaki, M., 2002. ADMIT: anomaly-based data mining for intrusions. In *ACM SIGKDD international conference on Knowledge discovery and data mining*. pp. 386–395. Available at: http://dl.acm.org/citation.cfm?id=775103.

Srikant, R. & Agrawal, R., 1996. *Mining sequential patterns: Generalizations and performance improvements*, Available at: http://link.springer.com/chapter/10.1007/BFb0014140 [Accessed March 29, 2014].

Swiss Federal Office of Energy, 2014. Energy Strategy 2050. Available at: http://www.bfe.admin.ch/themen/00526/00527/?lang=en [Accessed August 26, 2014].

Takeda, H. et al., 1990. Modeling Design Processes. *AI Magazine*, 11, pp.37–48. Available at: http://www.aaai.org/ojs/index.php/aimagazine/article/viewArticle/855.

Vaishnavi, V. & Kuechler, W., 2007. Design Research in Information Systems. *Order A Journal On The Theory Of Ordered Sets And Its Applications*, 48, pp.133–140. Available at: http://desrist.org/design-research-in-information-systems.

Villar, J., Cal, E. de la & Sedano, J., 2009. A fuzzy logic based efficient energy saving approach for domestic heating systems. *Integrated Computer-Aided Engineering*, 15, pp.1–9. Available at: http://iospress.metapress.com/index/L74647166223125U.pdf [Accessed May 22, 2014].

Ye, H., 2011. A Personalized Collaborative Filtering Recommendation Using Association Rules Mining and Self-Organizing Map. *Journal of Software*, 6(4), pp.732–739. Available at: http://ojs.academypublisher.com/index.php/jsw/article/view/4272 [Accessed January 31, 2015].

Youngblood, G., Cook, D. & Holder, L., 2005. A learning architecture for automating the intelligent environment. *PROCEEDINGS OF THE …*, pp.1576–1581. Available at: http://www.aaai.org/Papers/IAAI/2005/IAAI05-018.pdf [Accessed June 5, 2014].

Youngblood, G. & Heierman, E., 2005. Automation intelligence for the smart environment. ... *On Artificial Intelligence*. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.74.5893&rep=rep1&type=pdf [Accessed March 20, 2014].

FIGURES & TABLES

FIGURES

TABLES

GLOSSARY

AI	Artificial Intelligence
ALZ	Active LeZi
BAS	Building Automation Systems
CASAS	Continuous Adaptive Smart home Access System
COM	Continuous varied-Order Multi Threshold activity discovery
DVSM	Discontinuous Varied-Order Sequential Miner
ED	Episode Discovery
HMM	Hidden Marcov Models
HVAC	Heating, Ventilation and Air conditioning
IAAS	Infrastructure as a Service
MAV Home	Managing an Adaptive and Versatile Home
PUBS	Patterns of user behaviour system
RFID	Radio-frequency identification
RSYNC	Remote Synchronization Tool
SHIP	Smart Home Inhabitant Prediction
SMS	Short Message Service
SPEED	Sequence prediction via enhanced episode discovery
TEREDA	Temporal Relation Discovery of Daily Activities
VM	Virtual Machine
WSDD	Window Sliding with De-Duplication

APPENDIX

SMART HOME LOG FILE EXTRACT

[2014-12-30 18:01:26] ;StateApartment;zone.8.motion;2;inactive;;;;;Unknown;;

[2014-12-30 18:01:26] ;StateGroup;zone.8.light;2;inactive;Toilette;8;8;yellow;1;;

[2014-12-30 18:01:26] ;CallScene;Off-Automatic;40;Toilette;8;yellow;1;Unknown;0000000000000000000000000000000000;Scripting;

[2014-12-30 18:02:43] ;StateGroup;zone.11.light;2;inactive;Keller;11;11;yellow;1;;

[2014-12-30 18:02:43] ;CallSceneForced;Off-Automatic;40;Keller;11;yellow;1;Unknown;0000000000000000000000000000000000;Scripting;

[2014-12-30 18:03:09] ;BinaryInput;State 0;0;;;;;BM Garagem;302ed89f43f0000000000a8000013d5900;

[2014-12-30 18:03:09] ;StateDevice;dev.302ed89f43f0000000000a8000013d5900.0;2;inactive;;;;BM Garagem;302ed89f43f0000000000a8000013d5900;

[2014-12-30 18:03:09] ;StateApartment;zone.7542.motion;2;inactive;;;;;Unknown;;

[2014-12-30 18:03:09] ;CallSceneForced;Off;0;Garage;7542;yellow;1;Unknown;0000000000000000000000000000000000;Scripting;

[2014-12-30 18:03:34] ;CallScene;Off-Automatic;40;Treppe;14;yellow;1;Unknown;0000000000000000000000000000000000;Scripting;

[2014-12-30 18:03:43] ;CallSceneForced;Off;0;Keller;11;yellow;1;Unknown;0000000000000000000000000000000000;Scripting;

[2014-12-30 18:04:26] ;DeviceScene;Min;13;Toilette;8;Ventilador Banheiro EG;3504175fe0000000000000000006ef8b00;Unknown;;Scripting;

[2014-12-30 18:06:14] ;BinaryInput;State 1;0;;;;;BM Toilette;302ed89f43f0000000000a8000013d5200;

[2014-12-30 18:06:15] ;StateDevice;dev.302ed89f43f0000000000a8000013d5200.0;1;active;;;;BM Toilette;302ed89f43f0000000000a8000013d5200;

[2014-12-30 18:06:15] ;StateApartment;zone.8.motion;1;active;;;;;Unknown;;

[2014-12-30 18:06:15] ;StateGroup;zone.8.light;1;active;Toilette;8;8;yellow;1;;

[2014-12-30 18:06:15] ;CallScene;Stimmung1;5;Toilette;8;yellow;1;Unknown;0000000000000000000000000000000000;Scripting;

[2014-12-30 18:06:15] ;DeviceScene;Max;14;Toilette;8;Ventilador Banheiro EG;3504175fe0000000000000000006ef8b00;Unknown;;Scripting;

[2014-12-30 18:07:25] ;BinaryInput;State 0;0;;;;;BM Toilette;302ed89f43f0000000000a8000013d5200;

[2014-12-30 18:07:25] ;StateDevice;dev.302ed89f43f0000000000a8000013d5200.0;2;inactive;;;;BM Toilette;302ed89f43f0000000000a8000013d5200;

[2014-12-30 18:07:25] ;CallScene;Off-Automatic;40;Toilette;8;yellow;1;Unknown;0000000000000000000000000000000000;Scripting;

[2014-12-30 18:07:25] ;StateGroup;zone.8.light;2;inactive;Toilette;8;8;yellow;1;;

[2014-12-30 18:07:25] ;StateApartment;zone.8.motion;2;inactive;;;;;Unknown;;

[2014-12-30 18:10:25] ;DeviceScene;Min;13;Toilette;8;Ventilador Banheiro EG;3504175fe0000000000000000006ef8b00;Unknown;;Scripting;

[2014-12-30 18:12:51] ;StateGroup;zone.9.light;2;inactive;Küche;9;9;yellow;1;;

[2014-12-30 18:12:51] ;CallScene;Off;0;Küche;9;yellow;1;Luz Cozinha + Interruptor;3504175fe0000000000000000003140b00;dSM-API;

[2014-12-30 18:13:41] ;StateGroup;zone.2.light;1;active;Schlafzimmer;2;2;yellow;1;;

[2014-12-30 18:13:42] ;CallScene;Stimmung1;5;Schlafzimmer;2;yellow;1;Interruptor Carla e Danilo;3504175fe000000000000000000313cb00;dSM-API;

[2014-12-30 18:17:51] ;StateGroup;zone.5.light;1;active;Badezimmer;5;5;yellow;1;;

[2014-12-30 18:17:51] ;CallScene;Stimmung1;5;Badezimmer;5;yellow;1;Interruptor Banheiro Cima;3504175fe0000000000000000003136100;dSM-API;

[2014-12-30 18:18:03] ;StateGroup;zone.5.light;2;inactive;Badezimmer;5;5;yellow;1;;

[2014-12-30 18:18:03] ;CallScene;Off;0;Badezimmer;5;yellow;1;Interruptor Banheiro Cima;3504175fe0000000000000000003136100;dSM-API;

[2014-12-30 18:18:19] ;StateGroup;zone.9.light;1;active;Küche;9;9;yellow;1;;

[2014-12-30 18:18:19] ;CallScene;Stimmung1;5;Küche;9;yellow;1;Luz Cozinha + Interruptor;3504175fe0000000000000000003140b00;dSM-API;

[2014-12-30 18:24:36] ;BinaryInput;State 1;0;;;;;BM Eingang;302ed89f43f0000000000a8000013d5b00;

[2014-12-30 18:24:36] ;StateDevice;dev.302ed89f43f0000000000a8000013d5b00.0;1;active;;;;BM Eingang;302ed89f43f0000000000a8000013d5b00;

[2014-12-30 18:24:37] ;StateApartment;zone.7.motion;1;active;;;;;Unknown;;

[2014-12-30 18:24:37] ;StateGroup;zone.7.light;1;active;Eingang;7;7;yellow;1;;

[2014-12-30 18:24:37] ;CallSceneForced;Stimmung1;5;Eingang;7;yellow;1;Unknown;0000000000000000000000000000000000;Scripting;

[2014-12-30 18:24:46] ;CallScene;Bell;73;Broadcast;0;broadcast;0;Interruptor Campaínha;3504175fe00000000000000000009f2300;dSM-API;

[2014-12-30 18:25:43] ;BinaryInput;State 1;0;;;;;BM Toilette;302ed89f43f0000000000a8000013d5200;

[2014-12-30 18:25:43] ;StateDevice;dev.302ed89f43f0000000000a8000013d5200.0;1;active;;;;BM Toilette;302ed89f43f0000000000a8000013d5200;

[2014-12-30 18:25:43] ;StateApartment;zone.8.motion;1;active;;;;;Unknown;;

[2014-12-30 18:25:44] ;StateGroup;zone.8.light;1;active;Toilette;8;8;yellow;1;;

[2014-12-30 18:25:44] ;CallScene;Stimmung1;5;Toilette;8;yellow;1;Unknown;0000000000000000000000000000000000;Scripting;

[2014-12-30 18:25:44] ;DeviceScene;Max;14;Toilette;8;Ventilador Banheiro EG;3504175fe0000000000000000006ef8b00;Unknown;;Scripting;

[2014-12-30 18:25:45] ;StateGroup;zone.14.light;1;active;Treppe;14;14;yellow;1;;

[2014-12-30 18:25:45] ;CallScene;Stimmung1;5;Treppe;14;yellow;1;Luz Escada Keller + Interruptor EG;3504175fe000000000000000000313a300;dSM-API;

[2014-12-30 18:25:53] ;BinaryInput;State 0;0;;;;;BM Eingang;302ed89f43f0000000000a8000013d5b00;

[2014-12-30 18:25:53] ;StateDevice;dev.302ed89f43f0000000000a8000013d5b00.0;2;inactive;;;;BM Eingang;302ed89f43f0000000000a8000013d5b00;

[2014-12-30 18:25:53] ;StateApartment;zone.7.motion;2;inactive;;;;;Unknown;;

[2014-12-30 18:25:59] ;BinaryInput;State 1;0;;;;;BM Garagem Corredor;302ed89f43f0000000000a8000013d6c00;

[2014-12-30 18:26:00] ;StateDevice;dev.302ed89f43f0000000000a8000013d6c00.0;1;active;;;;BM Garagem Corredor;302ed89f43f0000000000a8000013d6c00;

[2014-12-30 18:26:00] ;StateApartment;zone.7542.motion;1;active;;;;;Unknown;;

[2014-12-30 18:26:00] ;CallScene;On-Area2;7;Garage;7542;yellow;1;Unknown;0000000000000000000000000000000000;Scripting;

[2014-12-30 18:26:00] ;CallScene;On-Area1;6;Garage;7542;yellow;1;Unknown;0000000000000000000000000000000000;Scripting;

[2014-12-30 18:26:08] ;StateGroup;zone.7.light;2;inactive;Eingang;7;7;yellow;1;;

[2014-12-30 18:26:08] ;CallSceneForced;Off;0;Eingang;7;yellow;1;Unknown;0000000000000000000000000000000000;Scripting;

[2014-12-30 18:26:17] ;StateGroup;zone.5.light;1;active;Badezimmer;5;5;yellow;1;;

[2014-12-30 18:26:17] ;CallScene;Stimmung1;5;Badezimmer;5;yellow;1;Interruptor Banheiro Cima;3504175fe0000000000000000003136100;dSM-API;

[2014-12-30 18:26:43] ;BinaryInput;State 0;0;;;;;BM Toilette;302ed89f43f0000000000a8000013d5200;

[2014-12-30 18:26:43] ;StateDevice;dev.302ed89f43f0000000000a8000013d5200.0;2;inactive;;;;BM Toilette;302ed89f43f0000000000a8000013d5200;

[2014-12-30 18:26:43] ;StateApartment;zone.8.motion;2;inactive;;;;;Unknown;;

[2014-12-30 18:26:44] ;StateGroup;zone.8.light;2;inactive;Toilette;8;8;yellow;1;;

[2014-12-30 18:26:44] ;CallScene;Off-Automatic;40;Toilette;8;yellow;1;Unknown;0000000000000000000000000000000000;Scripting;

[2014-12-30 18:26:59] ;BinaryInput;State 0;0;;;;;BM Garagem Corredor;302ed89f43f0000000000a8000013d6c00;

GUIDANCE FOR EVALUATION PARTICIPANTS

How the system works:

Your historical smart home event data was mined by algorithms for reoccurring behavioural patterns. The patterns represent frequent and periodic activities which occur in your household. These patterns are analyzed by a recommender system and compared to the current behaviour in your house. The aim of the system is to generate and send recommendations which should help you to save energy.

Since the log files from your house have to be copied, parsed and analyzed, there will be a latency between the occurrence of the event and the recommendation on your device. For this reason, each recommendation is given a timestamp, which should allow you to answer them, even you have already fulfilled the recommended action by yourself.

Example scenario:

A possible scenario is, that you have forgotten to turn off the light in your bedroom. If the system detects this deviation from your standard behaviour, it will generate a recommendation, which will be sent via SMS to your mobile phone (from +41 76 601 30 71) and will have the following format:

```
Hi test user,

I would recommend to turn Off-Device Nachttischlampe (on 2014-11-16 23:19:16).

Is this recommendation useful?

Yes:http://snurl.com/29fgj3d?r=2663&c=4f=1
No: http://snurl.com/29fgj3d?r=2663&c=4f=0
```

How the evaluation will proceed:

If you receive a recommendation on your device, please decide whether the recommendation was useful or not by clicking on the corresponding link. A useful recommendation is defined as a recommendation "which will lower energy usage but not decrease comfort levels". Since the system will only recommend actions, which will lower energy usage, it is your part to decide if there are any losses in comfort.

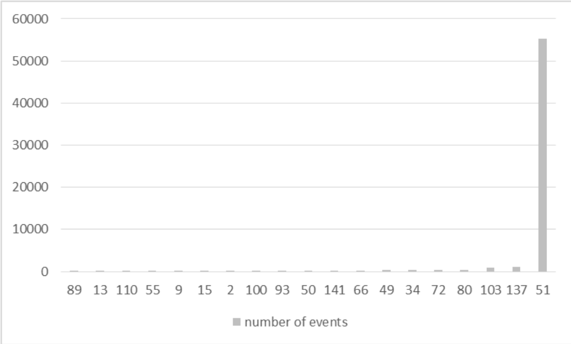
To keep the effort involved from your side to a minimum, your feedback is analyzed and considered for further recommendations. For example: if a recommendation generated from the same pattern was considered as not useful for ten times in a row, no further recommendations from this pattern will be sent to your mobile.

Contact detail:

If you have any questions or problems during the evaluation, please do not reply on recommendations but write to: michael.zehnder@students.fhnw.ch or phone number.

OUTLIER REMOVAL DESCRIPTION

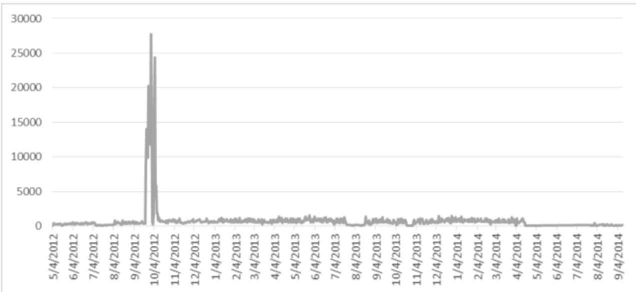
A closer look on the outliers showed that there was always one installation, which caused the unusual growth in event data points.



Almost all of the 55306 events could be assigned to one single trigger (device) id, which was not referenced in the database. To remove such inconsistent database entries, the references where checked and all which could not be assigned to a trigger (device) where removed from the database.

In another case, an event with the trigger “scripting” and the scene “Stimmung 1” was fired over 18000 times within four hours. This data could be attributed as testing and was deleted from the dataset.

All other outliers could be attributed to one installation, installation number 137. An app_button_click was called over 100’000 times from a device called “Hz TKM Mischer zu” within a few days.



The events were removed.

Furthermore, it was detected that events where the source device was “NULL” or “Unknown” where were often outliers. All these events were deleted, because they are useless for the system. The inhabitant would not understand a recommendation with the content “Please turn of Device Unknown”.