UNIVERSITY OF AMSTERDAM

Faculty of Science
Master Artificial Intelligence
Track Intelligent Systems

Behavior Analysis of Elderly using Topic Models

Master Thesis

Kristin Rieping

Supervisors:
Dr.ir. B.J.A. Krse
Dr. G. Englebienne
Dr. P.H. Rodenburg

Amsterdam, August 2013

Abstract

In this thesis two novel variations of the Latent Dirichlet Allocation (LDA) model are presented. The models give the opportunity to detect patterns in multi-dimensional data in an unsupervised manner. LDA-Gaussian is a combination of a Gaussian Mixture Model and a LDA model. Here the multinomial distribution of the topics, that is normally used in the LDA model, is replaced by a set of Gaussian Distributions. In this way similar looking sensor data is automatically grouped together and captured in the same topic. LDA-Poisson, the second variation of the model, takes a set of Poisson Distribution for the topic descriptions. This distribution makes it possible to handle discrete multi-dimensional data. The parameters of both models are determined with an EM-algorithm. Both models are applied to real sensor data, which is gathered in the homes of elderly people. It is shown that meaningful topics can be found and that a semantic description of these topics can be given.

Contents

Contents **v**

List of Figures **vii**

1 Introduction **1**

2 Related Work **3**

3 Data Description **5**
- 3.1 Homes and persons . . . 5
- 3.2 Sensors . . . 6
- 3.3 Received Data . . . 6

4 Features **9**

5 Topic Models **11**
- 5.1 Introduction to Topic Models . . . 11
- 5.2 Topic models with Sensor Data . . . 12
- 5.3 LDA-Gaussian Model . . . 13
  - 5.3.1 Motivation and Assumptions of the Model . . . 13
  - 5.3.2 Model Description . . . 14
  - 5.3.3 Variational Inference . . . 15
- 5.4 LDA-Poisson Model . . . 17

6 Experiments **19**
  - 6.0.1 Qualitative Results . . . 19
  - 6.0.2 Quantitative Results . . . 26

7 Future Work **29**

8 Conclusions **31**

Bibliography **33**

List of Figures

3.1 Floorplan of the houses with sensor descriptions 6
3.2 Sensor Data for two different hours at a day of one person. The fields 'kitchen','living room','bathroom','bedroom','hallway' are marked with the colors 'yellow','blue','red','green','purple' respectively. 7

4.1 Vector representation of the data. The data of the sensors is shown in the left image. It is translated in the vector shown on the right-hand side. 9

5.1 Graphical representation of the LDA model 12
5.2 The Bag-of-words model with topic assignment. 13
5.3 Feature space with k-means and topic assignment. 13
5.4 Gaussian Mixture Model with LDA 13
5.5 The two different topics are marked with red and blue. 13
5.6 Graphical representation of the LDA-Gaussian model 15
5.7 Approximation of the model. 16
5.8 Graphical representation of the LDA-Poisson model. 18

6.1 Topic distribution for 50 days for the Bag-of-Words model 20
6.2 Topic distribution for 50 days for the k-means model 20
6.3 20
6.4 Topic distribution for 50 days 20
6.5 Visualization of the topics 20
6.6 Topic distribution per day and the Topic visualization for LDA-Poisson 20
6.7 Topic distribution for 10 days 21
6.8 Visualization of the topics 21
6.9 Topic distribution per day and the Topic visualization for LDA-Gaussian 21
6.10 Topic distribution per day and the Topic visualization for LDA-Gaussian. HouseNr=1 21
6.11 Topic distribution per day and the Topic visualization for LDA-Poisson. HouseNr=1 21
6.12 Topic distribution per day and the Topic visualization for LDA-Gaussian. HouseNr=2 22
6.13 Topic distribution per day and the Topic visualization for LDA-Poisson. HouseNr=2 22
6.14 Topic distribution per day and the Topic visualization for LDA-Gaussian. HouseNr=3 22
6.15 Topic distribution per day and the Topic visualization for LDA-Poisson. HouseNr=3 23
6.16 Topic distribution per day and the Topic visualization for LDA-Gaussian. HouseNr=4 23
6.17 Topic distribution per day and the Topic visualization for LDA-Poisson. HouseNr=4 23
6.18 Topic distribution per day and the Topic visualization for LDA-Gaussian. HouseNr=5 24
6.19 Topic distribution per day and the Topic visualization for LDA-Poisson. HouseNr=5 24
6.20 Perplexity for different initializations of Topics for 5 different House 26
6.21 Perplexity for LDA-Gaussian and LDA-Poisson with different amount of time-slices 27
6.22 Perplexity for LDA-Gaussian and LDA-Poisson with different amount of topics . . 27
6.23 27
6.24 Perplexity (x-axis) for the hold-out-set for different amount of time-slices (y-axis) . 27

Chapter 1

Introduction

The life expectancy of people is assumed to rise continuously [11]. That means that also the percentage of elderly increases. Elderly people often need more health care and studies show that they like to live at home [4] . That is why monitoring the health condition of people in their home environment becomes more and more important.

New techniques give the possibility to monitor elderly from the distance or even automatically. These systems give the possibility to detect accidents, if for example a person falls on the ground [9], or monitor the health condition on a long term basis(find cite). Some of these techniques use cameras that are placed in the homes of elderly [10]. But these methods are privacy-sensitive and often not adopted by the elderly(cite).
Less intrusive methods use simple sensors like motion sensors or pressure mats that are placed in the homes of people [Tapia]. Reading and interpreting this sensor data is difficult and that is why activity recognition is often done to make the data more easy to interpret. Different techniques attempt to extract 'Activities of daily living' (ADL's). Changes in the ADL's can be a sign for declines of peoples health. (cite)
Some researchers depend on annotated data to find ADL's in sensor data [Tapia,Ontologies,Particle Filter]. But the task of labeling the data is time consuming and also might effect the output of the sensor data. If for example a person is asked to notate his activities during the day, the action of annotating also takes time and the sensor data might be inaccurate. A camera that is placed in the home environment might also affect the way people behave and the recorded data might also not be correct representation of peoples behavior.

Topic models seem to be a good way to find behavior patterns in data automatically without the need of annotation. Verschillende aanpakken zijn gebruikt om uit sensor data behavior patronen te vinden en de gevonden topics are comparable to ADL's. LDA is often used as a basis [Blei]. This model searches Latent topics in data, but this model is in eerste instantie devolped for clustering van text documenten. The Bag-of-words model (BOW), which is the basis for LDA, is not that easily appliable on sensor data. In the BOW model a dictionary is build of all possible words, which can be seen as independent dimensions. So for text document the data space that is used is a high dimensional binary space. Sensor data contains darentegen vaak not a lot of dimension, but the dimension is mostly not binary and the observation per dimension are more variate. Als is sensor data often depending on the time and there with the volgorde van eventen is also very important.
Different researchers apply the BOW to sensor data by discretizing the data in some way and are building virtual words of the observations. Often a way of preprocessing the data is neccessary to find the Vocabulary and apply the LDA model. In [Huynh,Cavallo] the words that are extracted from the sensor data are clustered with k-means to group variations in the sensor together and in this way reduce the size of the dictonary. In [Chikhaoui] sequential patterns are found in the data which are used as words in the dictonary. is here used for describing different text documents.

Sensor data is not so easily described with this method, due to the time dependency of the data and the small amount of dimensions which are though more variate in the values. With an unsupervised method labeling the data is not necessary. And systems can record for a long time and generate a lot of data. Verifying a method is in this way much harder but the patterns that are found with theses methods might give interesting results that can be used by care takers and health professionals.

In this work an algorithm is developed that does not need preprocessing of the data but can handle variations in the different dimensions of the data. The clustering is added into the LDA model itself. And a topic will be described with a set of parameters for a Gaussian distribution in every dimension. A variation of the algorithm is given where every dimension is described with a Poisson distribution. The algorithms are applied to real-life sensor data, that is gathered in the houses of solitary living elderly. The latent topics that are found describe the sensor data of a day, where a topic can be seen as an abstract description of an activity or a combination of activities. Example of such a topics are 'preparing breakfast' or 'going to toilet'.

Topic models are often used in the field of classifying text documents. The idea is that every document might belong to a couple of different topics and a topic can be described with a distribution over words. A newly seen document can then be assigned to some topics according to the words that occur in the document. In this way a document is described with a distribution of topics. A detailed description of the method and how the model parameters are found are described in the work of Blei et al. [1]. This method does not need any annotation of the data and finds the topics automatically.

To make use of this method a suitable feature representation of the available data is required. The features are found by counting the activations of sensor groups. In this way a relative low number of dimensions are found, which can contain positive integers. In theory these numbers have an infinite range, but in practice the count of sensor activations of every dimension is reduced to a maximum.

This feature representation differs a lot to the Bag-of-words model that is often used for the description of textual data. In this representation a dictionary of all possible words is build and a document can be described with a binary vector of the same length as the size of the dictionary. With respect to our data, text data thus has much more dimension but every dimension only contains a binary value.

Due to the feature representation of the sensor data the Bag-of-words model is not suitable. Therefore a model based on LDA is build that models every dimension with a Gaussian distribution. In this way small variations of the sensor data are captured. A variation of this model is build where the underlying distribution of the dimensions becomes the Poisson distribution. This suits the way the data is described better, because every dimension of the features contains a count of events. The parameters of both models are found with an EM-procedure, which uses the likelihood of the model to converge to the optimal model parameters.

In the next chapter of this thesis an overview of related approaches are given. In section **??** the data that is used is described in more detail and after that representation of the features is described in **??**. This is followed by section **??** which describes the different ways of the usage of Topic Models with the available data. The section **??** contains the different experiments that are performed. Section **??** the outcome of the experiments are discussed and finaly the conclusion of the report is given in section **??**.

Chapter 2

Related Work

Monitoring the health of elderly can be done in many different ways. Also the goal of monitoring elderly differs. One goal can be detecting accidents where in other systems the focus lays on the long-term health monitoring, which is in fact a way to prevent accidents. In [12] for example a system is described that is installed at peoples home and monitors their health. They use a variety of sensors and the system can give a detailed information of the health condition of the inhabitant. Sensors were added in bathtub, toilet and bed. These devices where specially build for the experiments which makes this approach quite expensive.
In [8] more simple sensors are used to monitor the health of solitary living people. Different sensors in the houses are tracking the health condition and the system is able to automatically contact emergency services or give the signal to send a caretaker at the home of the subject. A combination of static and wearable sensors is used, some of these sensor record automatically and other needs be triggered by the inhabitant.
Systems that monitor the health of people need to be not intrusive, which means that the choice of sensors becomes important. People do not want to be watched all the time, which makes the use of camera or microphones ([10], [15]) mostly not eligible. It is also desirable that sensors are low in cost. Another thing that is desirable is that the system collects data automatically. Inhabitants do not want to need to interact with the system constantly and the system should not affect their daily living. To meet these conditions several researchers focus on the task of activity recognition with simple sensors.

In [7] a system to recognize activities is introduced that is based on ontolgies. Simple binary sensors, like movement sensors, contact switches and pressure mats, are used. Activities are described with ontolgies that are build from the sensor data. In [13] contact switches that also generate binary data are used. A naive Baysian Network is trained with annotated data. This data is is collected from the inhabitants of two different houses. In the work of Wilson et al. [14] a particle filter is applied to binary sensor data. They focus on handeling multiple inhabitants in house.

All the work that is described above about activity recognition uses a supervised manner to detect activities in the data. Annotating the data has some disadvantages as it is pointed out in the work of Chikhaoui et al. [5]. In their work they use the topic model LDA in combination with sequential pattern mining. These sequential patterns build the words, that are needed as input for the LDA model. They apply their model to different data sets with different sensor types. The focus in this work lays more on recognizing activities than finding behavior pattern on a day as it is done in this thesis.
In the work of Castanedo et al. [3] an overview of different applications of the LDA model is given. They apply the LDA model to a data set collected from a sensor network in an office environment. The big amount of sensors installed in the test environment and big amount of data collected makes it possible to apply the Bag-of-words model to this data set. However they have difficulties to give an interpretation of the detected topics.

In the work of Casale et al. [2] LDA is apllied to data coming from a wearable sensor. They reduce the size of the dictionary by clustering the sensor readings. The size of the clusters is difficult to choose, which is also shown in this thesis. A different kind of data is used in the work of Farrahi et al. [6]. Cell information of mobile phones give information about the location of a person. With LDA they are able to find different kind of behaviors of people. They use a nice feature representation with which transitions of people can be modelled.

Chapter 3

Data Description

In this section we describe in what kind of houses the data is gathered and what kind of persons live there. We describe the sensors that are used and give an impression of the data that is received from these sensors.

3.1 Homes and persons

In the homes of five different people sensors are installed. The floor-plan of these homes is for all residents the same and is shown in figure 3.1. There might be small differences of the locations of the sensors due to the personal arrangements of peoples stuff. The persons that live in the homes are people that need healthcare on a regular basis, they are further able to live on their own. The amount of data collected differs for the different houses, but there is at least 63 days of data available for every house.

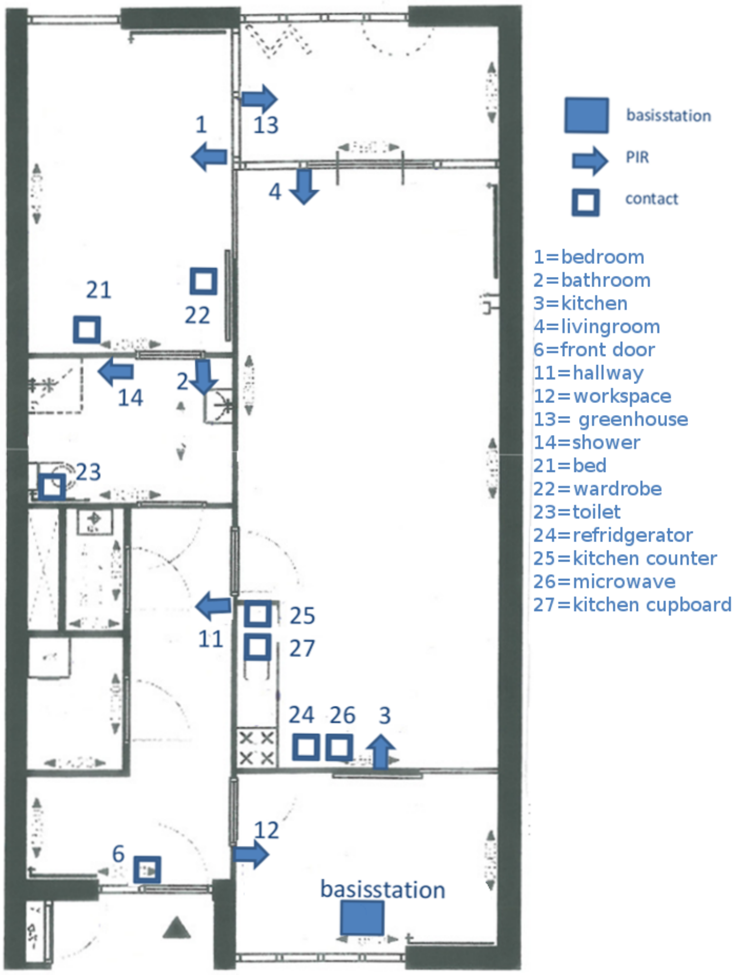


Figure 3.1: Floorplan of the houses with sensor descriptions

3.2 Sensors

There are different types of sensors installed in the homes. The contact switches are mostly installed at doors and cupboards. They get the value 'one' if a door is opened and the value 'zero' if the door is closed again. The motion-sensor (PIR) are placed at different places in the homes, mostly against the walls. They have a range of 5 meters. If a motion occurs in the region the sensor sends an impulse value, which means that the value becomes 'one' and immediately 'zero' again . After that the sensor is set to mute for about 3 minutes, which means that in this time there is no motion captured. In this way constantly firing of the sensor will be avoided. Every activity of the sensors is send to the basis-station 3 times, so that the chance is reduced that the data is not received. The sensor system is active 24 hours and 7 days a week. However failure can occur due to network problems, sensor failing or other unexpected problems.

3.3 Received Data

In figure 3.2 the data stream of two different hours of one day is shown. The data belongs to one person. Several sensors that are located in the same room are manually grouped together in a field. The fields are {'kitchen', 'living room', 'bathroom', 'bedroom', 'hallway'}. They are marked in the figure with different colors.

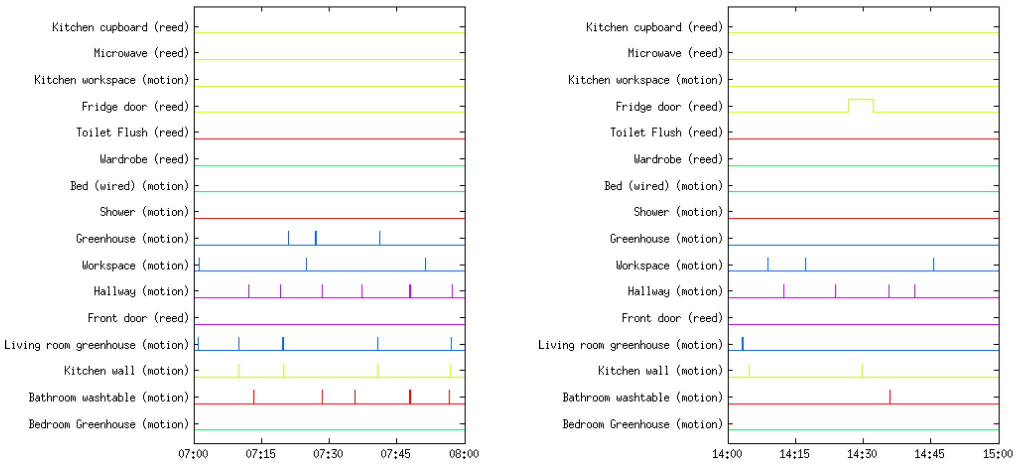


Figure 3.2: Sensor Data for two different hours at a day of one person. The fields 'kitchen','living room','bathroom','bedroom','hallway' are marked with the colors 'yellow','blue','red','green','purple' respectively.

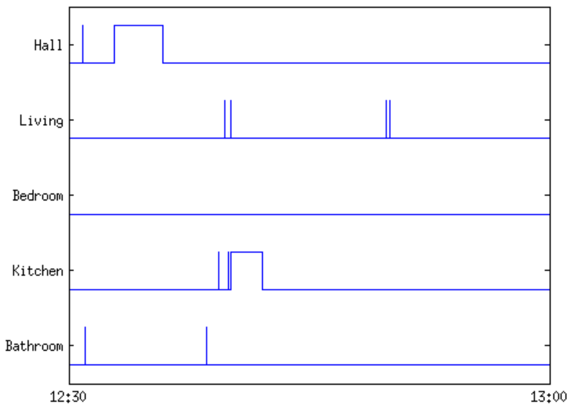
In the figure you can see the different type of data that is generated by the different sensor types. The fridge sensor is a reed sensor which has the value '1' for a longer period of time, when the door is opened for a while. The motion sensors on the other hand only give a impulse value as mentioned before. You can also see that some sensors are not triggered at all in the given time intervals.

Chapter 4

Features

The data that is received from the sensors generates a continuous data stream for every sensor. A good feature representation of the data is required so that the LDA model can be applied. First the data is divided into five fields and all the sensors in one field are grouped together. In this way the data to is reduced to five dimension. The continuous data stream cannot be used as input for the LDA model. That is why the data is divided into time-slices of length l. If for example the length of the time-slices is $l = 30$ min., the number of time-slices on one day is $n = 48$. A day starts and ends at 3 a.m. in the morning. In this way the chance to cut between activities is reduced. It still can occur that a person goes to bed late or that he needs to visit the toilet. For now this fact is left out in the part of modeling.

For every time-slice the number of sensor activations of one field are counted. Every field thus builds a dimension of the observations o_n. The length of time that a sensor stays in the active state is not taken into account. This is done because sometimes unrealistic high values are measured. These values can occur if someone leaves a door open of a cupboard for example. Then the observations contains a high value but does not contain a lot of information about the behavior. In figure ?? an example on how the data is translated into a vector representation is given for one time-slice.



$$\begin{bmatrix} o_{Hall} \\ o_{Living} \\ o_{Bedroom} \\ o_{Kitchen} \\ o_{Bathroom} \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 0 \\ 3 \\ 2 \end{bmatrix}$$

Figure 4.1: Vector representation of the data. The data of the sensors is shown in the left image. It is translated in the vector shown on the right-hand side.

An additional dimension for the time is used. There are two different ways how the time dimension is added to the observations. The fine-grain representation adds the number of the time-slice in which an observations is captured, at the end of the observation vector. In the coarse-grain representation the 24 hours of a day are divided into the five time intervals $\{3am - 8am, 8am - 1pm, 1pm - 18pm, 18pm - 23pm, 23pm - 3am\}$. So the observation of figure ?? will become $o_n = \{2, 4, 0, 3, 2, 2\}$ in the coarse-grain representation. The observation falls into the second time interval. In the fine-grain representation the observation will be $o_n = \{2, 4, 0, 3, 2, 20\}$ if the total number of time-slices on a day is $n = 48$.

The given feature representation, with the two variation of the time dimensions, are used in the following chapters. But there are plenty more ways to describe the data so that it can be used in the LDA model. For example the size of the time-slices can be changed. This leads to a higher resolution and may give more detailed information of the behavior. Another way to get a different feature representation is to combine sequential time-slices as it is described in the work of Farrahi et al. [6]. So for a given time-slice add the observation values of the previous and the subsequent time-slice. This will lead to a 15 dimensional observation plus one dimension for the time value. In this way the transitions are taken more into account, which might contain valuable information over the behavior.

We also might want to combine different sizes of time-slices into one observations, so that the global and detailed view is combined.

The different kind of sensor are also maybe important. The reed sensor, which is mostly installed at doors might contain more important information than the motion sensor. The motion sensor is also triggered by small movements, whereelse if a door is opened from a cupboard you can assume that an important action has taken place, as for example the person is grabbing a plate to prepare a meal. So it might be an idea to give a higher weight to sensor activities of reed sensors in house.

It is obvious that the feature space can be made nearly infinity big and it is quite a challenge to find the best feature representation. The likelihood that is gained from the EM-algorithms might be a good indication for a good feature representation.

Chapter 5

Topic Models

In the introduction of this section we describe the general idea of topic models followed by a section that describes how this kind of models can be used on the sensor data that we have. After that we introduce the extension of the LDA model which combines the clustering and topic estimation in one algorithm. And after that LDA-Poisson model is described.

5.1 Introduction to Topic Models

Topic models are often used in the field of document classification. Given a set of documents (corpus) it is assumed that every document belongs to one or more topic(s). So for example a news article may belong for some percentage, let us say 30 %, to the topic 'Economy' and for 70 % to the topic 'Politics'. Another document of the same Corpus may belong to the topic 'Economy' with 50 %, 'Politics' with 30 % and 'Global Warming' with 20 %. There might be a lot of different topics and the topics can have different level of details.

The topics are defined by several words that can occur in the documents. The topic 'Economy' may be defined by the list of words {'trade', 'industry','GDP'}. Other topics have different lists of words that describe them. The list might be longer or shorter and the words in the list will depend on the corpus that is used to generate the topics. It might also be the case that one word belongs to multiple topics. Eventually we can find the topic distribution of a document according to the words that are included in this document.

In the topic model 'Latent Dirichlet Allocation' (LDA) it is assumed that a Corpus can be made out of a generative process. The parameters that generate the corpus are than used to describe the model of the corpus. The generative process which builds a corpus is as follows:

For every document that is generated in the Corpus

1. Choose the amount of words in the document from $N \sim Possoin(\xi)$.
2. Choose a topic distribution $\theta \sim Dir(\alpha)$ for the document.
3. For each of the N words w_n:
   (a) Choose a topic $z_n \sim Multinomial(\theta)$.
   (b) Choose a set of words w_n from the set of all words V from $p(o_n|z_n;\beta)$, a multinomial probability conditioned on the topic z_n. Where β is the distribution over words given a topic.

The model is also shown in figure 5.1. The parameters α and β define a corpus.

To determine the model parameters an EM-algorithm can be used. How this algorithm can be applied is extensively described in [1].

In the next section we describe how this model can be used with the sensor data, that is gained in the different houses.

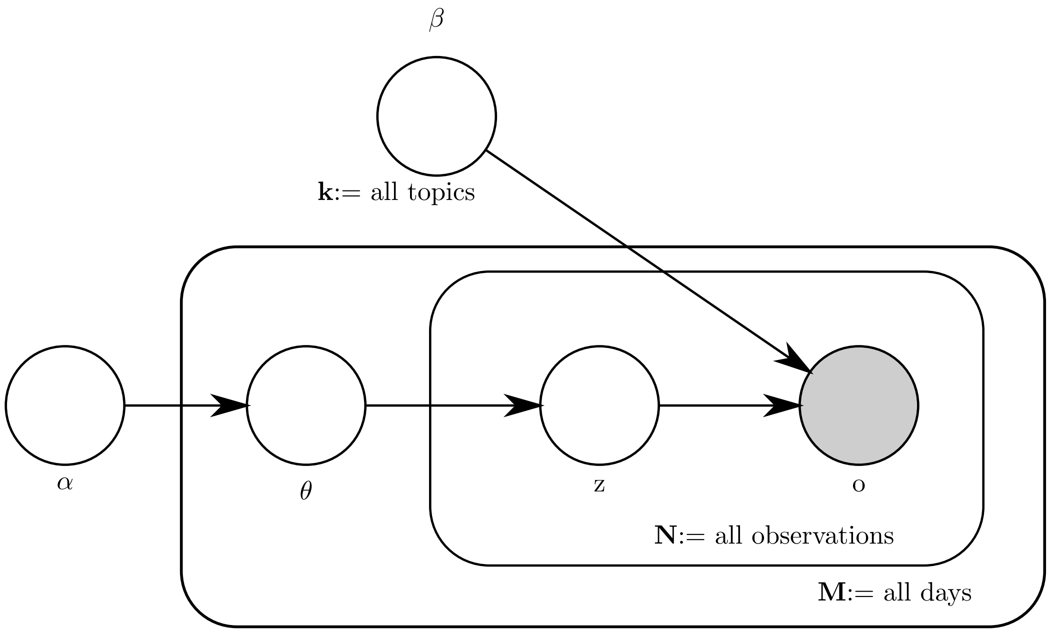


Figure 5.1: Graphical representation of the LDA model

5.2 Topic models with Sensor Data

In order to employ the topic model with the sensor data we first introduce the different levels of description given our data and relate them to the terms of document classification.

- **Dataset/Corpus**: One dataset C describes the sensor data that is gained in the home of a single person. So for every person there is a separate dataset. This set of data can be compared with a Corpus in document classification.
- **Day/Document**: Every Dataset is divided in days. A day can be compared with one document in a Corpus.
- **Observations/Words**: Finally every day is build of a set of observations. The amount and dimension of the observations depend on the representation of the features, which is described later in chapter **??**. Observations can be roughly compared with words in document classification.

There are some differences between the data that is used for topic detection in documents and our sensor data.
The main difference is that words that look similar to each other, like "illusion" and "allusion", may belong to a totally different topic in the document classification. But in our case, two observation that are similar to each other, are more likely to refer to the same topic. So for example if the topic "preparing food" has the observation using fridge 3 times in it, an observation of using fridge 4 times may also refer to the same topic. That is why we cannot directly compare the words in a text document with the observations used in the sensor data.
Another difference is that in the topic model, LDA, it is assumed that the order of the words, in that they appear in the text, does not matter. In our case the time when an observation is made is of big influence. We can overcome this problem by adding time as an additional dimension to our observations.

If the Bag-of-Words model will be applied to the data, a dictionary of all unique observations must be made. All this observations then can be seen as a different dimension which are independent of each other. In figure 5.2 it is shown how the BOW is used. Observations that are assigned to the same topic do not need to lie close to each other in the feature space and the correct topics might not be found with LDA. This approach needs a lot of data to find meaningful results. The size of the 'dictionary' varies depending on the feature representation.

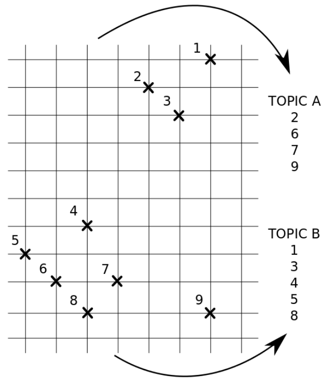


Figure 5.2: The Bag-of-words model with topic assignment.

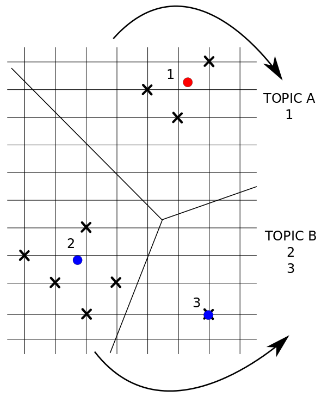


Figure 5.3: Feature space with k-means and topic assignment.

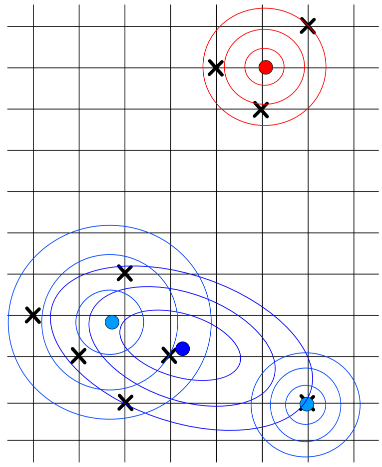


Figure 5.4: Gaussian Mixture Model with LDA

Figure 5.5: The two different topics are marked with red and blue.

To make sure that similar observations will be grouped together and will appear in the same topic, we can apply the k-means algorithm to find clusters in the data. The id's of the centroids of the clusters then function as 'words' and can be used in the same way as Bag-of-words model, but then with reduced dictionary size. So if we apply LDA to the clustered data some clusters may fall into the same topic. This is shown in figure 5.3.
The outcome of the LDA model does strongly depend on the outcome of the apriori used cluster algorithm. The clusters that are found with k-means are maybe not a good representation of the data. All clusters may have the same amount of influence on the LDA model, which might not be the correct way to describe the data. The Gaussian Mixture model might give a better description of the clusters. Combining the part of clustering directly into the topic model will avoid the apriori clustering and the clusters are then formed due to the topics, which will give a better globalization of the data. In figure 5.4 we show how the combination of the part of clustering into the Topic Model might improve the Gaussian Mixture Model.
According to the feature representation, which in our case is a discrete derivation of the sensor data, a Poison distribution might be a good choice to model the topic distributions. The Poisson distribution can be used instead of the Gaussian distribution.

5.3 LDA-Gaussian Model

In this section we explain how the part of clustering is combined within the LDA model itself. Instead of the k-means clusters we use a Gaussian distribution to model every dimension within the features.

5.3.1 Motivation and Assumptions of the Model

Using the cluster algorithm k-mean in advance has some disadvantages. First of all we need to choose the number of clusters. This is not always that easy and then it is still not assured that the best clusters are found. Every cluster also has a hard separation line, which does not give any degradation if a value is far away from the mean of the cluster. There is no quality measurement of the clusters given. So every cluster is equally probable to occur in the topic model, which is not always desirable.

With a Gaussian Mixture Model we might be able to distinguish between more or less important clusters. But we might want to let the topic model decide which topics have more influence and which have not, according to the data.
That is why we combined the clustering part and the topic estimation into one step. Instead of a multinomial distribution over all observations that occur in the 'Dictionary', we model a Gaussian distribution over every dimension of the observations. In this way similar observation can be captured in the model itself and are not generalized in one cluster beforehand.
The difference is that instead of a large, fixed set of unique observations (Dictionary), with a multinomial distribution, we take a Gaussian distribution over every dimension of the observations. In this way smoothing is not necessary, because unseen observations will be handled properly. In the next sections we describe the model in more detail. We first give an overview of the generative process. Then we explain the variational inference that is necessary to make the parameter estimation possible. And finally explain the EM-algorithm that determines the parameters.

5.3.2 Model Description

Our model assumes that every day in a dataset can be represented as random mixtures of latent topics, where every topic can be described as a distribution over observations. We assume the following generative procss for every day m in the data set C:

1. The amount of observations on a day m is fixed with size N (for every day the same size).
2. A day has a distribution over the topics given with $\theta \sim Dir(\alpha)$.
3. For each of the N observations on a day o_n:
   (a) Estimate the topic $z_n \sim Multinomial(\theta)$.
   (b) An observation o_n is gained from $p(o_n|z_n, \boldsymbol{\mu}, \boldsymbol{\sigma})$, which is a probability that can be drawn from a set of Gaussian distributions. This probability is conditioned on the topic z_n and the Gaussian Parameters $\vec{\mu_i}$ and $\vec{\sigma_i}$ of length d that belong to the estimated topic i.

In this model the amount of topics k is assumed to be known and fixed and with it the size of the topic variable z. The probability for the observations is parametrized with two matrices $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, both of size $D \times k$, where D is the amount of dimensions in an observation and k the amount of topics. They present the mean and standard deviation respectively and for every topic i and every dimension d there is a set of parameters, which describes a Gaussian distribution. Every value of a dimension for an observation o_{ndi} can then be drawn from a Gaussian Distribution $\mathcal{N}(\mu_{di}, \sigma_{di})$. The size of the dimension D is assumed to be fixed and a more extensive description of the representation of the observations is given in section **??**. α represents the Dirichlet parameter and is vector of length k.

In figure 5.6 the graphical representation of the model is shown. It differs from the basic LDA model (figure 5.1) in the description of the topic distribution β, which is here replaced with μ and σ.

Assuming that the generative process described above can describe the available data properly, we now want to find the parameters so that the model best describes our data. We want in fact maximize the probability for the dataset C given the model with respect to the parameters α, μ and σ. This probability looks like this

$$p(C|\alpha, \mu, \sigma) = \prod_{m=1}^{M} p(m|\alpha, \mu, \sigma) \tag{5.1}$$

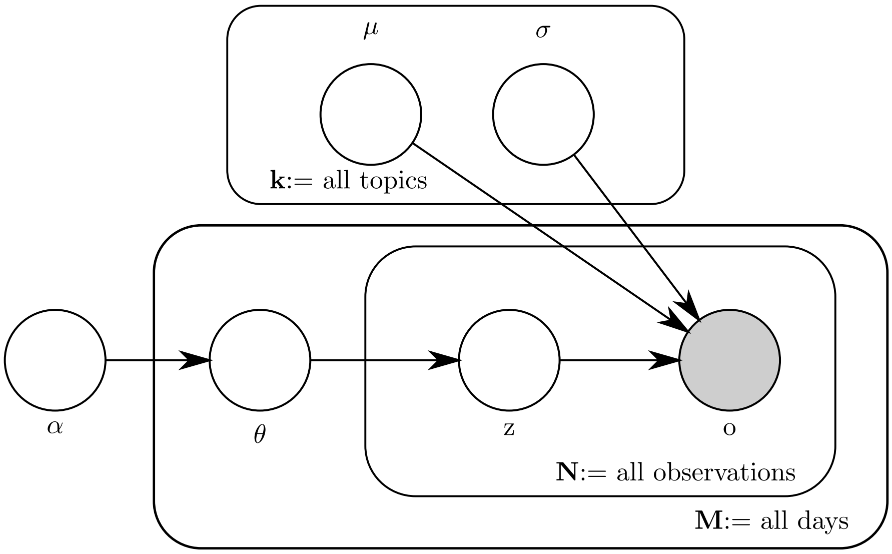


Figure 5.6: Graphical representation of the LDA-Gaussian model

where M is the amount of days within the dataset and the probability for a day m given the three parameters, which is the marginal distribution over a day, is

$$p(m|\alpha, \mu, \sigma) = \int p(\theta|\alpha) \left(\prod_{n=1}^{N} \sum_{z_n} p(z_n|\theta) p(w_n|z_n, \mu, \sigma) \right) d\theta \tag{5.2}$$

This distribution is gained by integrating the joint distribution 5.3 by θ.

$$p(\theta, \mathbf{z}, \mathbf{w}|\alpha, \mu, \sigma) = p(\theta|\alpha) \prod_{n=1}^{N} p(z_n|\theta) p(w_n|z_n, \mu, \sigma) \tag{5.3}$$

In the next section we describe how we can find the optimal parameters for the model given a data set.

5.3.3 Variational Inference

The marginal distribution, which is given in the previous section, can be written in terms of the parameter α, μ and σ as

$$p(m|\alpha, \mu, \sigma) = \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \int \left(\prod_{i=1}^{k} \theta_i^{\alpha_i - 1} \right) \left(\prod_{n=1}^{N} \sum_{i=1}^{k} \prod_{d=1}^{D} \theta_i \mathcal{N}(o_{nd}, \mu_{id}, \sigma_{id}) \right) \tag{5.4}$$

Due to the coupling between θ and the Gaussian parameters μ and σ this probability is intractable to compute.

That is why we use a convexity-based variational algorithm to approximate the log-likelihood of a given dataset. An approximation of the model is given with

$$q(\theta, z|\gamma, \phi) = q(\theta|\gamma) \prod_{n=1}^{N} q(z_n|\phi_n). \tag{5.5}$$

In this model γ represents the Dirichlet parameter and ϕ are the multinominal parameter which can be viewed as the probabilty $p(z_i|o_n)$ and is given as a $k \times N$-matrix for every day m. The graphical representation of the model is shown in figure 5.7.

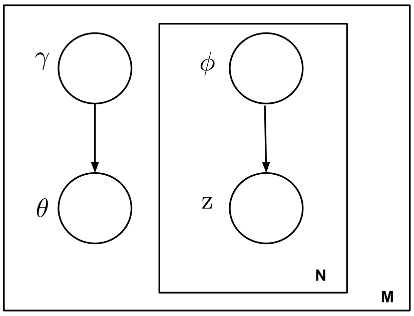


Figure 5.7: Approximation of the model.

Given the variational distribution for aproximate model we can estimate the lower bound of the log-likelihood with the Jensen inequality as

$$\begin{aligned} L(\gamma;\phi;\alpha;\mu;\sigma) =& E_q[\log p(\theta|\alpha)] + E_q[\log p(\mathbf{z}|\theta)] + E_q[\log p(\mathbf{w}|\mathbf{z},\mu,\sigma)] \\ & - E_q[\log p(\theta)] - E_q[\log q(\mathbf{z})] \end{aligned} \tag{5.6}$$

In terms of the model parameters and the variational parameters this becomes

$$\begin{aligned} L(\gamma;\phi;\alpha;\mu;\sigma) =& \log\Gamma(\sum_{j=1}^{k}\alpha_j) - \sum_{i=1}^{k}\log\Gamma(\alpha_i) + \sum_{i=1}^{k}(\alpha_i - 1)(\Psi(\gamma_i) - \Psi(\sum_{j=1}^{k}\gamma_j)) \\ & + \sum_{n=1}^{N}\sum_{i=1}^{k}\phi_{ni}(\Psi(\gamma_i) - \Psi(\sum_{j=1}^{k}\gamma_i)) \\ & + \sum_{n=1}^{N}\sum_{i=1}^{k}\sum_{d=1}^{D}\phi_{ni}\log(\mathcal{N}(o_{nd};\mu_{id},\sigma_{id})) \\ & - \log\Gamma(\sum_{j=1}^{k}\gamma_j) + \sum_{i=1}^{k}\log\Gamma(\gamma_i) - \sum_{i=1}^{k}(\gamma_i - 1)(\Psi(\gamma_i) - \Psi(\sum_{j=1}^{k}\gamma_j)) \\ & - \sum_{n=1}^{N}\sum_{i=1}^{k}\phi_{ni}\log\phi_{ni} \end{aligned} \tag{5.7}$$

With an EM process we are then able to maximize this lower bound on the log-likelihood. The two steps are:

1. **E-step:** For each day m, optimize the variational parameters $\{\gamma_m*, \phi_m*\}$
2. **M-step:** Maximize the resulting lower bound on the log-likelihood with respect to the model parameters α, μ and σ.

We now give a more detailed description on both of these steps.

E-step In the e-step of the algorithm the variational parameters ϕ and γ are optimized. To get the update function for ϕ we get all terms of the lower bound of the loglikelihood in equation (5.7) that contains the variable ϕ. Take $y_i = \sum_{d=1}^{D}\mathcal{N}(o_{nd};\mu_{id},\sigma_{id})$ We add the constraint $\sum_{i=1}^{k}\phi_{ni} = 1$ to the formula and get

$$L_{[\phi_{ni}]} = \phi_{ni}(\Psi(\gamma_i) - \Psi(\sum_{j=1}^{k}\gamma_j)) + \phi_{ni}\log(y_i) + \lambda_n(\sum_{j=1}^{k}\phi_{ni} - 1) \tag{5.8}$$

From this equation we take the derivative of the formula and set it to zero. This leads to the first update function

$$\phi_{ni} \propto \log(y_i) \exp(\Psi(\gamma_i) - \Psi(\sum_{j=1}^{k} \gamma_j)) \tag{5.9}$$

For γ we also take all terms of equation 5.7 that contain this variable and set the derivate to zero. This leads to the second update equation

$$\gamma_i = \alpha_i + \sum_{n=1}^{N} \phi_{ni} \tag{5.10}$$

M-Step In the m-step the parameters of the Gaussian distribution μ and σ are estimated with the weighted arithmetic mean calculated over all observation in a dataset given the parameter ϕ, which is gained in the previously e-step. This leads to the update formulas

$$\mu_{di} = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N} o_{dn} \phi_{ni}}{\sum_{m=1}^{M} \sum_{n=1}^{N} \phi_{ni}} \tag{5.11}$$

and

$$\sigma_{di} = \sqrt{\frac{\sum_{m=1}^{M} \sum_{n=1}^{N} o_{dn}^2 \phi_{ni}}{\sum_{m=1}^{M} \sum_{n=1}^{N} \phi_{ni}} - \mu_{di}^2} \tag{5.12}$$

To calculate the parameter α we again take all terms of the likelihood that contains the variable α. The derivative in the Hessian form is

$$\frac{\partial L}{\partial \alpha_i \alpha_j} = m(i,j) M \Psi'(\alpha_i) - \Psi'(\sum_{j=1}^{k} \alpha_j) \tag{5.13}$$

On this equation we can use the Newton-Rhapson method to calculate the optimal α.

5.4 LDA-Poisson Model

If the features are described in a discrete way, which means that every dimension has an integer value, the Poisson distributions is a better choice to describe the data. In figure 5.8 the graphical representation of LDA-Poisson model is shown.

The variational Inference and the EM-procedure will be the same as described before, except for the Gaussian distribution that is exchanged with the Poisson distribution. The parameter *lambda* which describes the Poisson distribution is calculated in the M-step with

$$\lambda_{di} = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N} o_{dn} \phi_{ni}}{\sum_{m=1}^{M} \sum_{n=1}^{N} \phi_{ni}}. \tag{5.14}$$

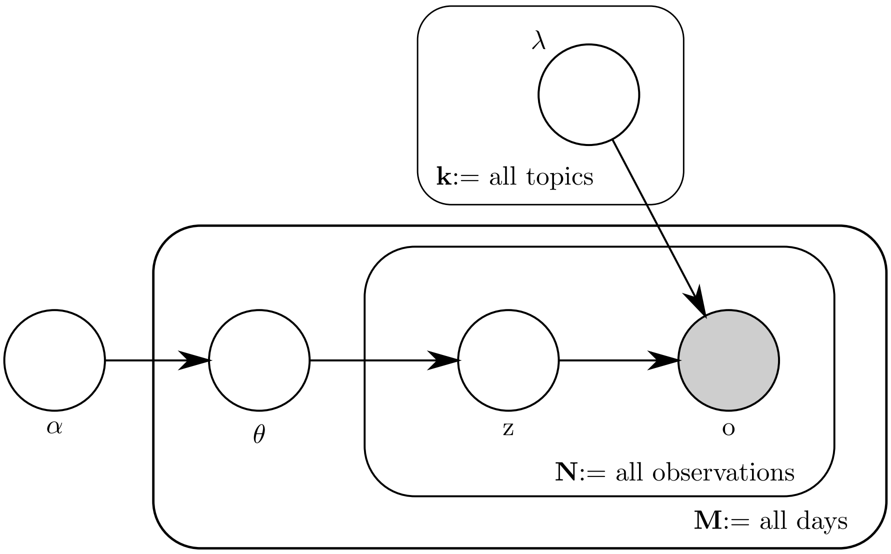


Figure 5.8: Graphical representation of the LDA-Poisson model.

Chapter 6

Experiments

In this section we describe the experiments that are performed and show some results. In the section of the qualitative results we give some visualizations of the topics that are found with the different models. We show that the topics can be meaningful and that semantic descriptions can be given to some of the topics.
In the section of the quantitative results we compare the two developed models with each other and show the performance for different number of topics and different number of time-slices.

6.0.1 Qualitative Results

In the following some visualizations of the different models are shown. First the topic distributions over some days are represented, for the different models. There are four model compared with each other. The first two models are the LDA model with a Bag-of-word representation with and without pre-clustering the dictionary with k-means. The other two model are the LDA-Gaussian model and the LDA-Poisson model which are described above.
After that a more detailed view of the topics is given and some semantic meanings of the topics are described.

For the visualization of the topics every time-slice is marked with a different color. The color depends on the topic that the time-slice belongs to. The belonging topic is the one with the highest probability of the observation in the given time-slice.

Comparison of the different models

In the following figures we used $N = 96$ time-slices for a day. The time dimension has a coarse-grain representation of the time, which means that there are five different time values. The number of topics that are initialized are $k = 5$. With the given representation the LDA model with a BOW-representation is still able to find different topics in the data, although the BOW representation without the clustering only distinguish between two topics (see figure 6.1). If the data is clustered on forehand LDA is able to find four different topics in the data (see figure 6.2). The number of clusters for the k-means algorithms is set to $V = 6$.

In figure 6.6 the outcome of the LDA-Poisson model is shown with the same variable values as before. On the right-hand side of the figure the topic description of the five topics that are found. The colors of the topics in the two images are equal to each other, so every red time slice on the left image is the topic that is shown with red in the right image. The six dimension of the observations are shown in this figure (6.5) on the x-axis. On the y-axis of every subfigure the λ-value of the Poisson distribution is shown. You can see (figure 6.4) that the LDA-Poisson model gives a relative high priority on the time value, because every day is separated into three timezones (yellow,green,red) and only the two other topics (blue, purple) are generated with respect of the sensor values of the five fields.

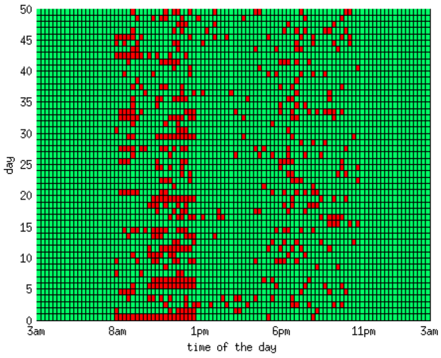


Figure 6.1: Topic distribution for 50 days for the Bag-of-Words model

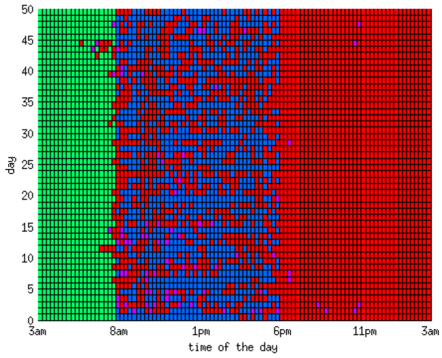


Figure 6.2: Topic distribution for 50 days for the k-means model

Figure 6.3

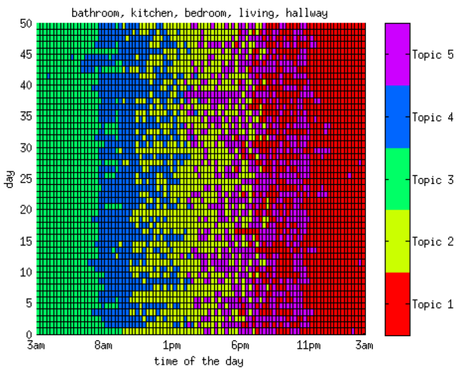


Figure 6.4: Topic distribution for 50 days

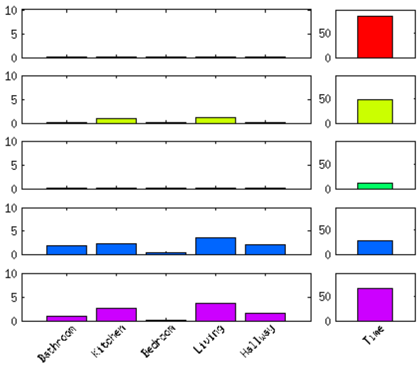


Figure 6.5: Visualization of the topics

Figure 6.6: Topic distribution per day and the Topic visualization for LDA-Poisson

Figure 6.9 shows the outcome of the LDA-Gaussian model again with the same variable values. In the right figure (6.8) on the x-axis again the 6 dimensions of the observations are shown. On the y-axis the mean value μ of the Gaussian distribution is shown and in the bar-chart the standard deviation σ is given with a vertical black line. You can see that the red topic here has a high σ-value and captures all time-slice where no sensor data is measured. The purple topic seems to represent the 'Going to the toilet at night' topic. Here the mean value of the time is low and only a few sensor measurements are captured. You can see that this model captures more sophisticated topics, that represent more meaningful results.

An attempt for semantic topic description

In the following figures the topic distribution for the five different houses is shown for the two models, LDA-Gaussian and LDA-Poisson. The number of time-slices is set set to $N = 48$ and the models are initialized with $k = 20$ topics and the time has a fine-grain representation. For every house and model 50 days of the data is shown. The ten topics with highest value are shown, time-slices that belong to a different topic are marked with gray.

You can see that for some people the structure of their daily behavior is more easily found than for other people. And also the different model give slightly different results. For some people you can clearly see the topic 'Going to toilet at night'. For house number 4 this topic is marked

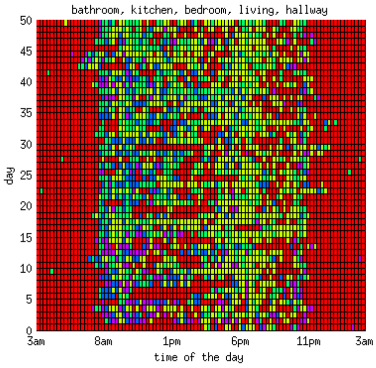


Figure 6.7: Topic distribution for 10 days

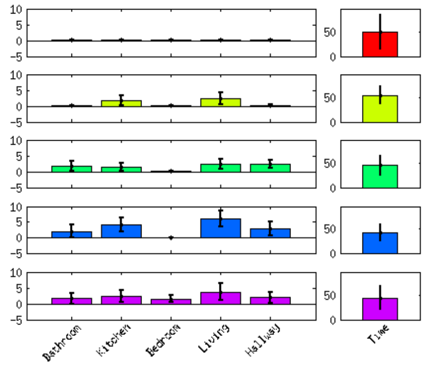


Figure 6.8: Visualization of the topics

Figure 6.9: Topic distribution per day and the Topic visualization for LDA-Gaussian

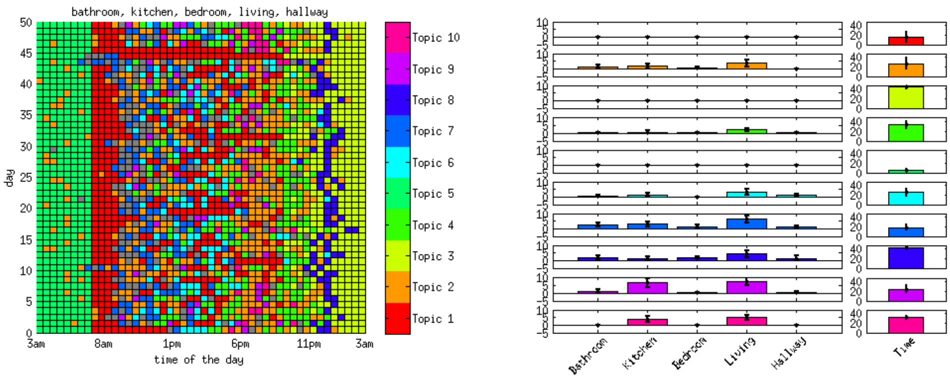


Figure 6.10: Topic distribution per day and the Topic visualization for LDA-Gaussian. HouseNr=1

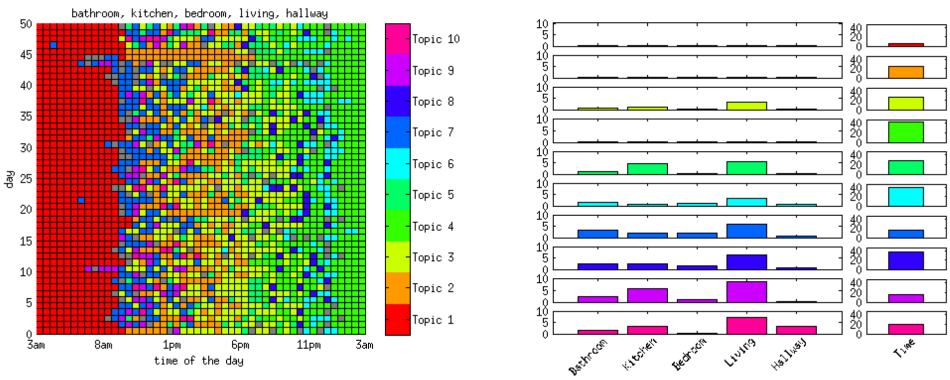


Figure 6.11: Topic distribution per day and the Topic visualization for LDA-Poisson. HouseNr=1

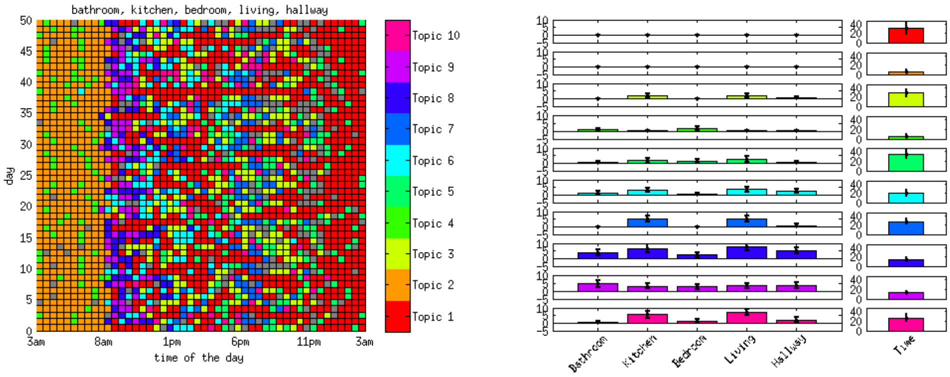


Figure 6.12: Topic distribution per day and the Topic visualization for LDA-Gaussian. HouseNr=2

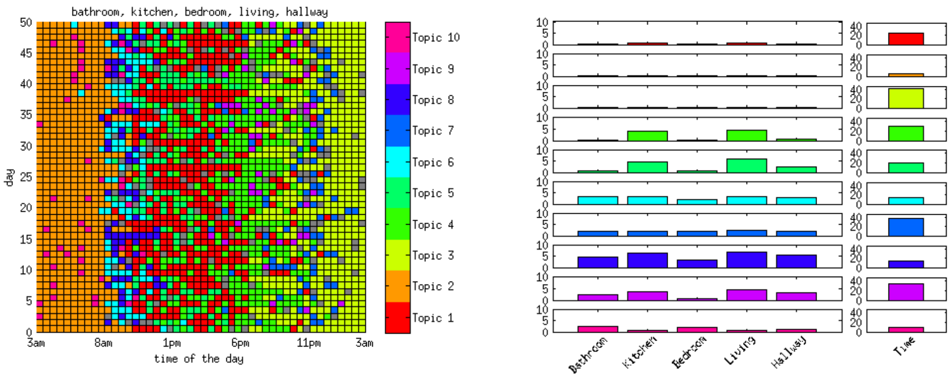


Figure 6.13: Topic distribution per day and the Topic visualization for LDA-Poisson. HouseNr=2

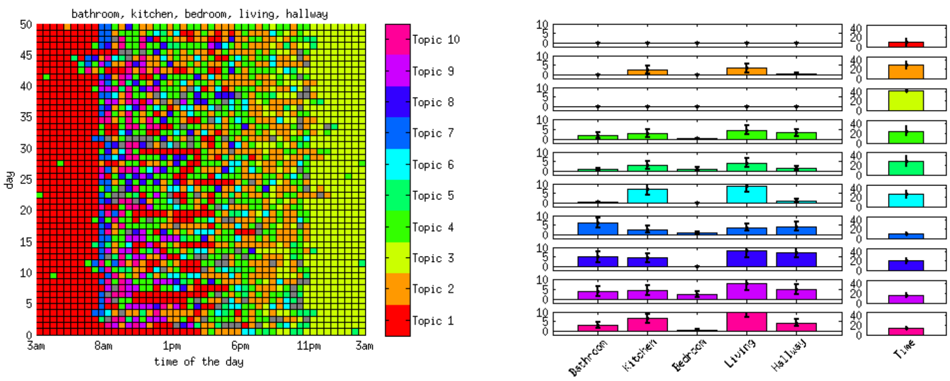


Figure 6.14: Topic distribution per day and the Topic visualization for LDA-Gaussian. HouseNr=3

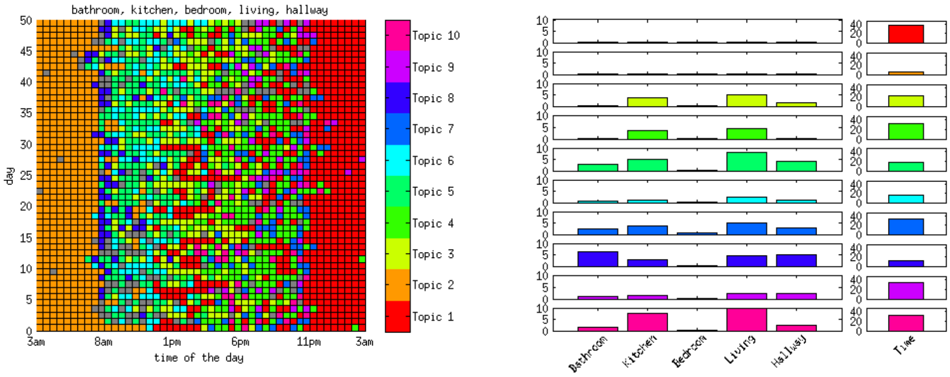


Figure 6.15: Topic distribution per day and the Topic visualization for LDA-Poisson. HouseNr=3

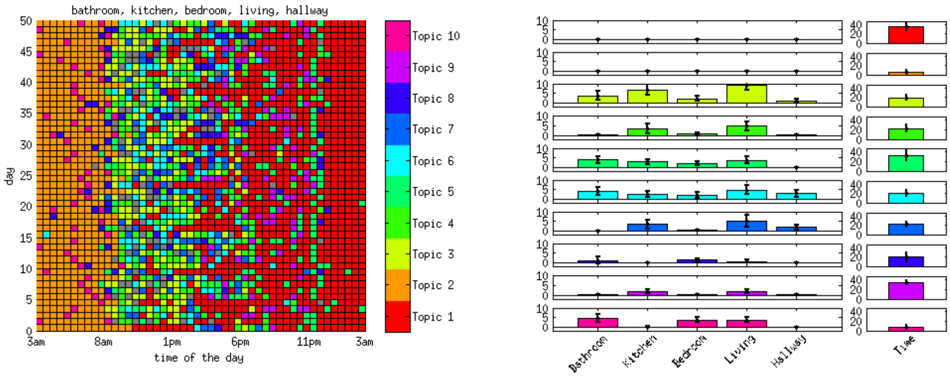


Figure 6.16: Topic distribution per day and the Topic visualization for LDA-Gaussian. HouseNr=4

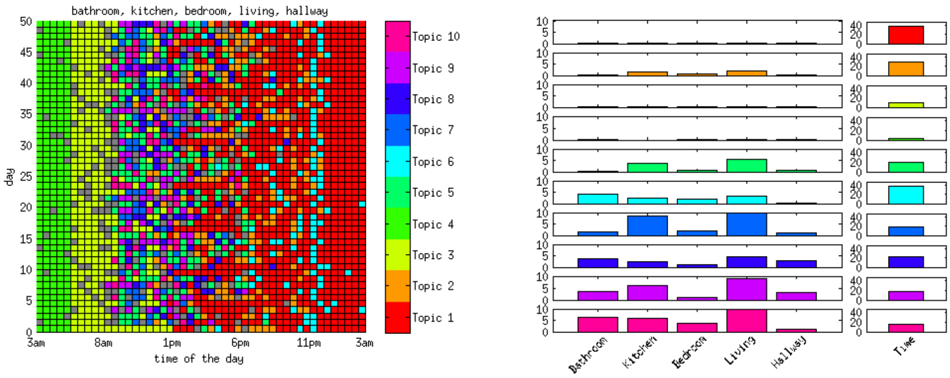


Figure 6.17: Topic distribution per day and the Topic visualization for LDA-Poisson. HouseNr=4

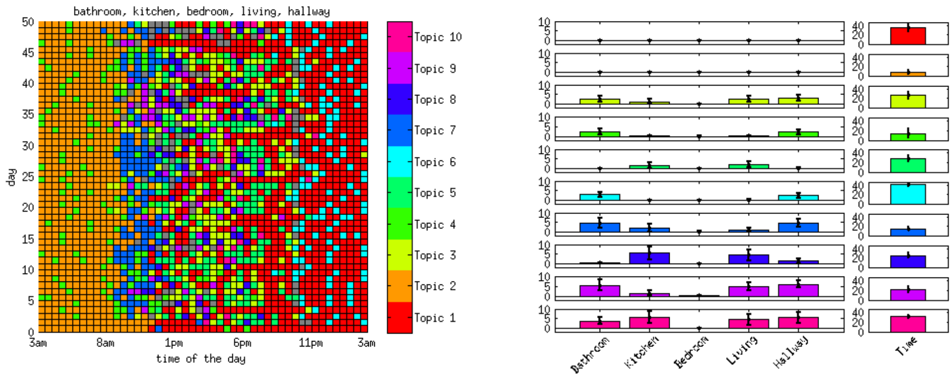


Figure 6.18: Topic distribution per day and the Topic visualization for LDA-Gaussian. HouseNr=5

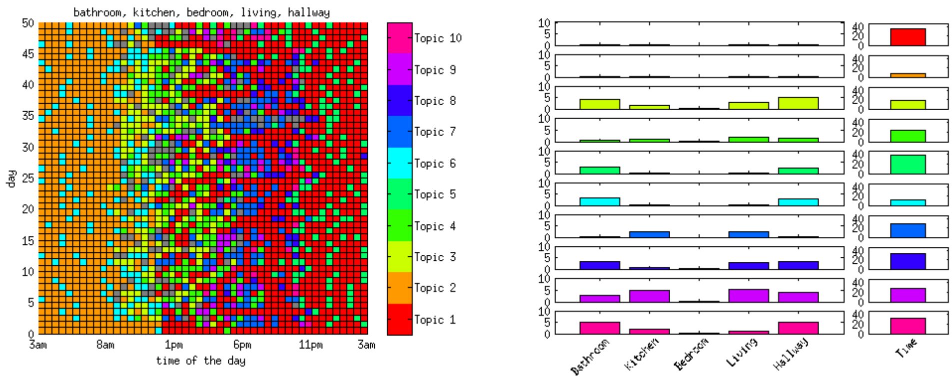


Figure 6.19: Topic distribution per day and the Topic visualization for LDA-Poisson. HouseNr=5

with pink in the LDA-Gaussian model. The three fields bathroom, bedroom and living are active and the mean time value is quiet low, which corresponds with an early time. For house number one this topic is also captured in the LDA Gaussian model, but not in the LDA-Poisson model.
You can also see the different behavior patterns between the persons. For example the person of house number 5 tends to sleep a little bit longer than the persons of other houses. Some behaviors are more regular than others. The LDA-Poisson model tends to focus on the time a little bit more then the LDA-Gaussian model.

6.0.2 Quantitative Results

To see if our results not only hold on our training data we we wish to also find a high log-likelihood for a hold-out set. In this way we make sure that the model does not over-fit the data. So for a given hold-out set we can calculate the perplexity with

$$perplexity(D_{HOS}) = exp\left\{-\frac{\sum_{m=1}^{M}\log p(\mathbf{o}_d)}{M * N}\right\} \tag{6.1}$$

In the next experiments we use 10% percent of our data as a hold-out set. We train the model on the rest of the data for different initialization values and calculate the perplexity for every run.

Different Sets of Data In figure 6.20 the perplexity is shown for the 5 different data-sets gained from the 5 houses. Every run is performed ten times and we took the mean over these runs for every initialization of amount of topics k. Every run is initialized with 5 days. The data sets vary in length, but as you can see in the figure, the amount of data is not necessary of influence how well the LDA-Gaussian model can be trained. Some data sets are much more stable than others and this is probably due to the way how regular peoples behavior is.

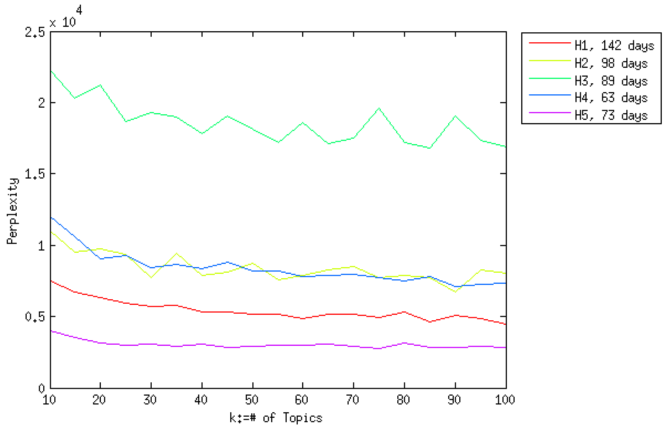
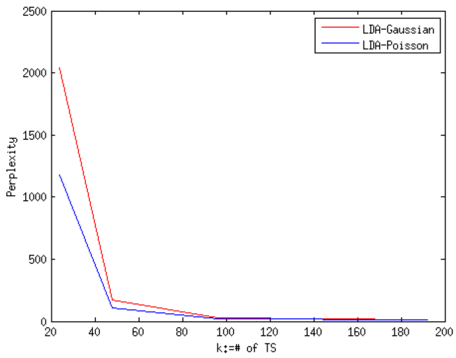


Figure 6.20: Perplexity for different initializations of Topics for 5 different House

Comparison LDA-Gaussian and LDA-Poisson To compare the two models with each other we drop the time dimensions in the observations and calculate the perplexity for the hold-out set with different amount of time-slices (figure 6.21) and with different amount of topics (figure 6.22). You can see that the LDA-Gaussian model outperforms the LDA-Poisson model for small amount of time-slices as well as for small amount of topics. When the amount of time-slices increases the performance of both models becomes similar to each other.

Different length of time-slices In figure 6.24 we give the perplexity for different length of time-slices for the 5 Houses. We take a again the mean over 10 runs. We take for every house the same amount of days to train and test the data. We can see that with more time-slices the perplexity becomes lower for the hold-out set. For house $H2$ and $H4$ the perplexity is much better for a small amount of time-slices. The perplexities for these two houses drop much faster if the amount of time-slices increases.



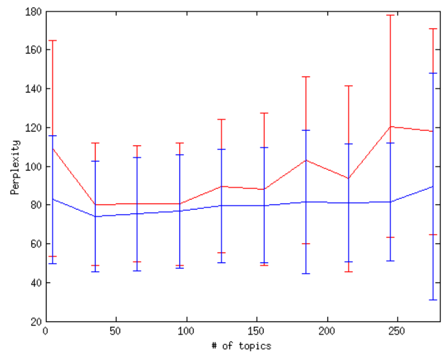


Figure 6.21: Perplexity for LDA-Gaussian and LDA-Poisson with different amount of time-slices

Figure 6.22: Perplexity for LDA-Gaussian and LDA-Poisson with different amount of topics

Figure 6.23

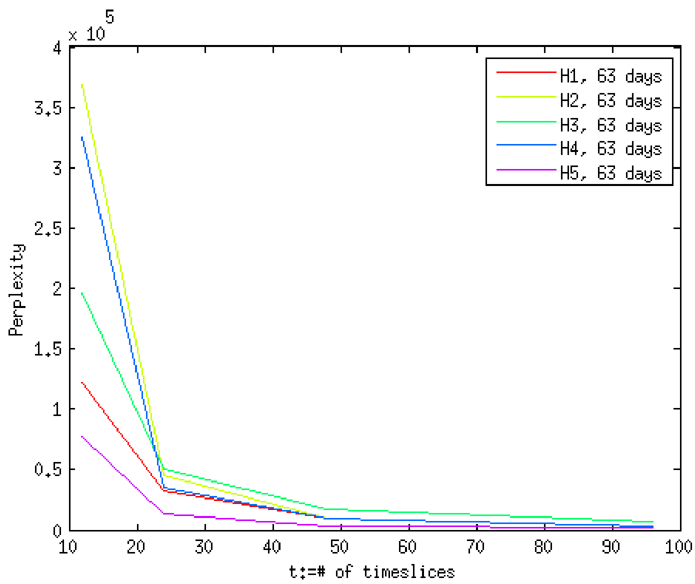


Figure 6.24: Perplexity (x-axis) for the hold-out-set for different amount of time-slices (y-axis)

Chapter 7

Future Work

Chapter 8

Conclusions

Write your conclusions here.

Bibliography

[1] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March 2003. 2, 11

[2] Pierluigi Casale, Oriol Pujol, Petia Radeva, and Jordi Vitrià. A first approach to activity recognition using topic models. In *Proceedings of the 2009 conference on Artificial Intelligence Research and Development: Proceedings of the 12th International Conference of the Catalan Association for Artificial Intelligence*, pages 74–82, Amsterdam, The Netherlands, The Netherlands, 2009. IOS Press. 4

[3] Federico Castanedo, Diego López de Ipiña, Hamid K. Aghajan, and Richard Kleihorst. Learning routines over long-term sensor data using topic models. *Expert Systems*, pages n/a–n/a, 2013. 3

[4] F Cavallo, G van den Broek, and C Wehrmann. *AALIANCE Ambient Assisted Living Roadmap.* Ambient Intelligence and Smart Environments. IOS Press, Amsterdam, 2010. 1

[5] Belkacem Chikhaoui, Shengrui Wang, and Hélène Pigot. Adr-splda: Activity discovery and recognition by combining sequential patterns and latent dirichlet allocation. *Pervasive and Mobile Computing*, 8(6):845–862, 2012. 3

[6] Katayoun Farrahi and Daniel Gatica-Perez. Daily routine classification from mobile phone data. In *Proceedings of the 5th international workshop on Machine Learning for Multimodal Interaction*, MLMI '08, pages 173–184, Berlin, Heidelberg, 2008. Springer-Verlag. 4, 10

[7] Xin Hong, Chris Nugent, Maurice Mulvenna, Sally McClean, Bryan Scotney, and Steven Devlin. Evidential fusion of sensor data for activity recognition in smart homes. *Pervasive and Mobile Computing*, 5(3):236 – 252, 2009. ¡ce:title¿Pervasive Health and Wellness Management¡/ce:title¿. 3

[8] Ohbyung Kwon, Jae Moon Shim, and Geunchan Lim. Single activity sensor-based ensemble analysis for health monitoring of solitary elderly people. *Expert Systems with Applications*, 39(5):5774 – 5783, 2012. 3

[9] Muhammad Mubashir, Ling Shao, and Luke Seed. A survey on fall detection: Principles and approaches. *Neurocomputing*, 100(0):144 – 152, 2013. ¡ce:title¿Special issue: Behaviours in video¡/ce:title¿. 1

[10] Norihiro Nagai, Yusaku Fujii, Hiroshi Ueda, Koichi Maru, and Shigeru Kumakura. Pc-based monitoring system for efficient use of day care centers for elderly people. *Procedia - Social and Behavioral Sciences*, 2(1):204 – 208, 2010. ¡ce:title¿The 1st International Conference on Security Camera Network, Privacy Protection and Community Safety 2009¡/ce:title¿. 1, 3

[11] United Nations. *World population to 2300.* United Nations, 2004. 1

[12] Toshiyo Tamura, Tatsuo Togawa, Mitsuhiro Ogawa, and Mikiko Yoda. Fully automated health monitoring system in the home. *Medical Engineering & Physics*, 20(8):573 – 579, 1998. 3

[13] Emmanuel Munguia Tapia, Stephen S. Intille, and Kent Larson. Activity recognition in the home using simple and ubiquitous sensors. In *In Pervasive*, pages 158–175, 2004. 3

[14] D. H. Wilson and C. Atkeson. Simultaneous tracking and activity recognition (star) using many anonymous, binary sensors. In *Proceedings of the Third international conference on Pervasive Computing*, PERVASIVE'05, pages 62–79, Berlin, Heidelberg, 2005. Springer-Verlag. 3

[15] Huadong Wu. *Sensor Fusion for Context-Aware Computing Using Dempster-Shafer Theory*. PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, December 2003. 3