UNIVERSITY OF AMSTERDAM

Faculty of Science
Master Artificial Intelligence
Track Intelligent Systems

Behavior Analysis of Elderly using Topic Models

Master Thesis of

Kristin Rieping

Supervisors:
Dr.ir. B.J.A. Kröse
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 low-dimensional sensor 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 low-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

List of Figures

Chapter 1

Introduction

The life expectancy of people is assumed to rise continuously in the following years [**?**]. As a consequence the percentage of elderly increases. Elderly people often need more health care but studies show that they prefer to live at home [4]. The manpower to take care of elderly people is not always available. That is why monitoring the health condition of people in their home environment becomes more and more important. In this way slowly emerging declines in the health condition of people can be detected and appropriate health care can be granted if it is necessary before critical health conditions are reached.

New systems give the possibility to monitor elderly from the distance or even automatically [12]. Different systems give the possibility to detect accidents or monitor the health condition on a long time basis. Some systems still need the action of a person involving into the system. For example the inhabitant has to push a button if an accident occurs [9]. Other systems use cameras to monitor the elderly [10, **?**]. But these methods are privacy-sensitive and often not adopted by the elderly.

Less intrusive methods use simple sensors like motion sensors or pressure mats that are placed in the homes of people [13, **?**]. Extracting valuable information out of this data is however difficult and that is why activity recognition is often done to make the data more easily 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 [13].
Some researchers depend on annotated data to find ADL's in sensor data [13, 7, 14]. But the task of labeling the data is time consuming and also intrusive, if the labeling is done by a different person. The knowledge of being observed by a sensor, for example a camera, might change the behavior patterns of a person and in this way the data is not accurate. Annotation that is done by the subject himself also might not by accurate, because the person has to interrupt his daily behavior to write down the annotation of the activities.

A way to automatically find behavior pattern in sensor data is by applying a topic model to the data. Topic models are initially designed for classifying text documents and they are able to find abstract topics, such as 'politics', 'sports', 'finances' etc. But a number of researchers have applied the topic model 'Latent Dirichlet Allocation' (LDA) to different kind of sensor data [6, 5]. The found topics are comparable to ADL's and can give a good representation of peoples daily behavior.
There is however a mayor difference between textual and sensor data. In the LDA model presented by Blei [1] the documents are represented by a Bag-of-words (BOW) model. A BOW model is an unordered representation of words, that does not take the grammar or the order of the words into account.
To make use of the LDA model on sensor data one has to create artificial words. But sensor data is mostly time dependent and that is why in most approaches a time value is added to the artificial

words. There are numerous ways ways to create words from sensor data.
Some researchers create artificial words by simply adding a time-value to the the sensor data and directly apply the BOW model to these 'words' [6, 3]. This approach requires a large dictionary of words to find behavior patterns in the data, to be able to capture all the variations in artificial words. For this reason other researchers first cluster the data and then apply the LDA model [8, 2]. In this way the size of the dictionary is smaller and less data is required. But finding the correct clusters is however difficult.

In this thesis the two new topic models 'LDA-Gaussian' and 'LDA-Poisson' are developed. These models are able to capture similar observations/words into the same topic automatically. The clustering and the LDA model are combined into one model and except for the construction of the artificial words, no further pre-processing is necessary .
In the original LDA model (Blei) the topics are described by a multinomial distribution over the words of the vocabulary. In the 'LDA-Gaussian' model this distribution is replaced with a Gaussian distribution, so that similar words are caught in the same topic. 'LDA-Poisson' is a variation on this model, where the underlying distribution is Poisson. In this way event based sensor data can be modeled.
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. The models are applied on real sensor data, that is obtained from the houses of solitary living elderly. All sensors are binary and are experienced as non-intrusive by the inhabitants. These elderly people live in a care home and get health care on a regular basis. Still they are able to live on their own.

In the next chapter of this thesis an overview of related approaches are given. In chapter 3 the data that is used is described in more detail and after that representation of the features is given in 4. This chapter is followed by chapter 5 which introduces the 'LDA-Gaussian' and 'LDA-Poisson' models. Chapter 6 contains the different experiments that are performed on the available data. In chapter 7 suggestions for future work are given and the conclusion of the thesis is presented in chapter 8.

Chapter 2

Related Work

The goal of monitoring the health of people is to detect accidents or even more important prevent accidents and critical health conditions. Changes in the daily behavior patterns can be a sign of changes in the health of people. This can be both mental or physical declines. There are different ways to monitor the health condition of people. Cameras or microphones can be very useful to monitor peoples behavior [10, 15]. But they are invading the privacy of people and often not accepted as sensors in peoples homes.

Simple binary sensors such us motion sensors, contact switches or pressure mats are preferable for health monitoring in home environments. These sensors are low in cost and easy to install. They are also often experienced as non-intrusive and not disturbing by the inhabitants. Numerous researchers implemented different approaches to apply activity recognition on this kind of data. These activities and especially changes in these activities, often referred to as ADL's, can then give valuable information on peoples health [**?**].
Tapia et al. [13] uses a naive Bayes classifier to find activities in annotated, sensor data. They show that it is possible to find activities in ubiquitous, simple sensor data, that was obtained in real-life environments.
In the work of Kasteren et al. [] two approaches for recognizing activities in sensor data are compared. The Hidden Markov Model and the Conditional Random Field are both applied to annotated, real-life sensor data. They also vary between different kind of sensor readings and show that this can improve the results for recognizing activities with their approaches.
Wilson et al. [14] implemented a Particle Filter to find activities in simulated as well as real-life data. They are able to distinguish the actions between multiple people in the environment.
Hong et al. [7] uses ontologies to describe daily activities. They use an evidential network to describe activities in a hierarchical way.

In the previous approaches annotated data is used to find activities in sensor data. Generating labeled data is however difficult. It is time-consuming and the labels can be inaccurate. Variations in the way people behave cannot be captured. That is why unsupervised methods to find activities or behavior patterns in sensor data is more convenient that supervised methods. Various authors applied LDA to different kind of data. This topic model is able to find abstract descriptions of activities in data automatically.
Chikhaoui et al. [5] uses the topic model LDA in combination with sequential pattern mining to find activities in various datasets. The sequential pattern are used as words that are needed as input for the LDA model. In their work they focus on detecting activities and not so much on daily behavior patterns of people.
Huyhn et al. [] and Casale et al. [] both apply LDA to sensor data obtained from wearable sensors. From acceleration features, that are clustered in advance, they generate the artificial words. The clustering is necessary to group similar words together. In this way the size of the dictionary is reduced and LDA can find meaningful topics in the data. Choosing the amount of clusters on

forehand is however difficult and has a big influence on the outcome of the LDA model.
Phung et al. [11] applys LDA to data that is gained of a WiFi network. They find behavior patterns of people in their work environment.
Farrahi et al. [6] applys LDA to location data gained from cell information of mobile phones. A lot of data is available and the simple location description that is used to create the artificial words make it possible to apply the BOW representation of the data directly. No clustering of the data is necessary.
In the work of Castanedo et al. [3] 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 also makes it possible to apply the BOW model to this data set directly. However they have difficulties to give an interpretation of the detected topics.

Chapter 3

Data Description

In this chapter a description of the houses and the inhabitants is given. The different kind of sensors are explained and an impression of the received data is given.

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 personal belongings. 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. An overview of the different houses is given in table 3.1

HouseNr	1	2	3	4	5
# of days	142	98	89	63	73
mean activities per day	488	523	668	565	427

Table 3.1

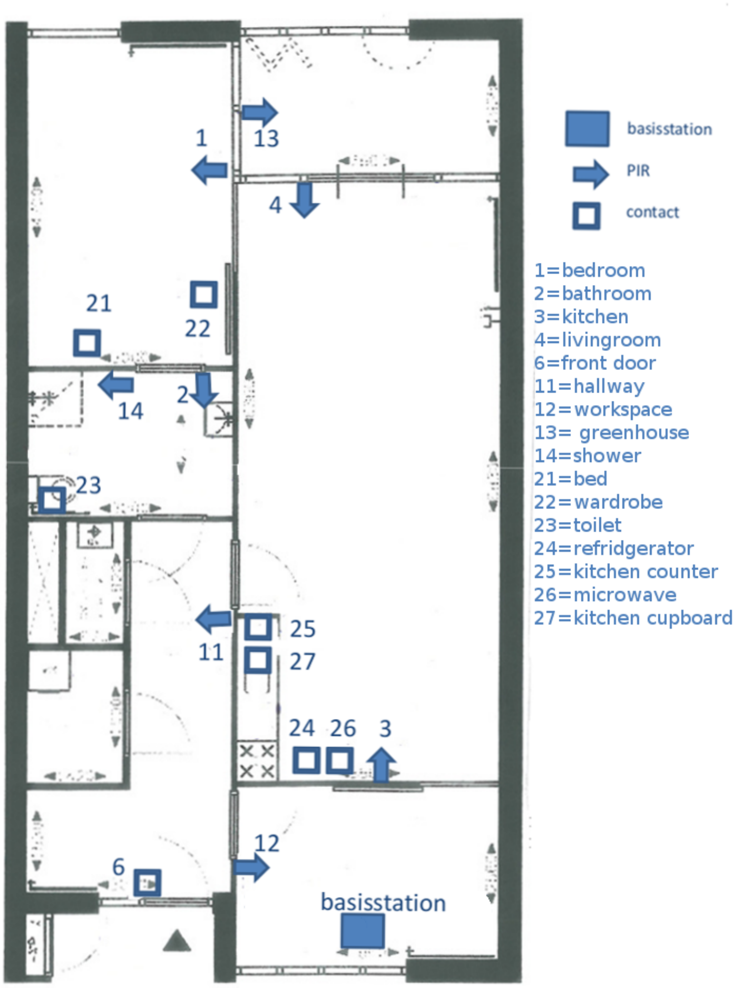


Figure 3.1: Floorplan of the houses with sensor descriptions

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. 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.

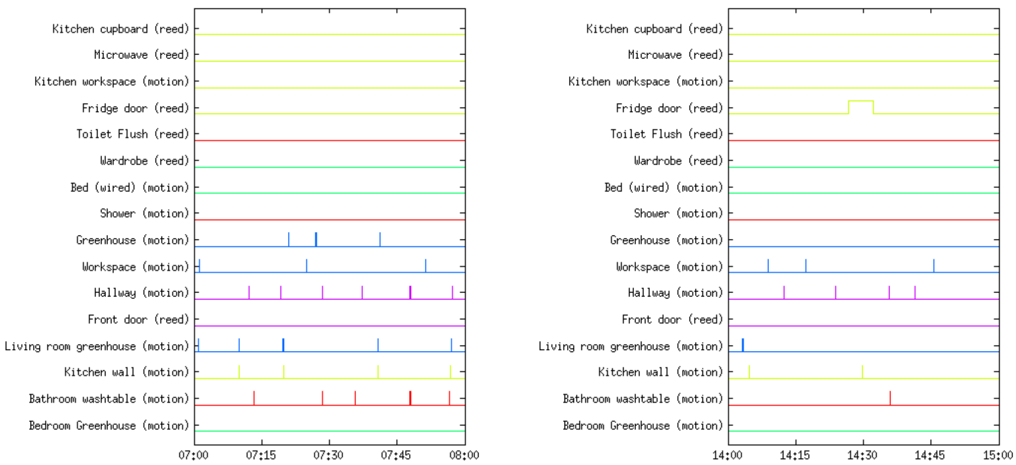


Figure 3.2: Sensor Data for two different hours at a day of one person. The fields 'kitchen','living room','bathroom','bedroom','hallway' are marked with the colors 'yellow', 'blue', 'red', 'green', 'purple' respectively.

In the figure one 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. Some sensors are not triggered at all in the time intervals that are shown.

Chapter 4

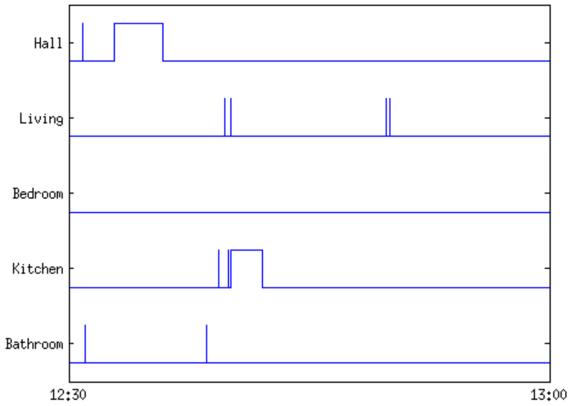
Features

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. The feature representation that is described in this chapter is used as the artificial words for the LDA model that is described in chapter 5 about the Topic model. First the data is divided into five fields and all the sensors in one field are grouped together. In this way the data is reduced to five dimensions. The continuous data stream cannot be used as input for the LDA model. Like it is done in the work of Farrahi [6] the data of one day 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 $n = 48$. For a chosen number of the time-slices n the length of time-slices for one day can then be calculated with $l = 1440/n$ in minutes. In the experiments (chapter 6) the number of time-slices is varied to see the effect on the topic model. 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. In one time-slice for each sensor field the number of times is counted that the signal changes from zero to one. The duration of an active signal, thus for how long a sensor field has the value one, is not taken into account. This is done because for example a door that accidentally is left open would otherwise generate a high value although this value is not very informative and would though disturb the data. Then the observations contains a high value but does not contain a lot of information about the behavior. Every sensor field then builds a dimension of the observations o_n. In figure ?? an example on how the data is translated into a vector representation is given for one time-slice.

The last dimension of the observation o_n represents the time value of the given time-slice. 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, which is also used in the work Farrahi [] and Castanedo [], 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 $n = 48$.

If the number of time-slices is set to $n = 48$ the maximal value that is observed in one field is 28. This is an extreme value and occurs not that often in the data. If the maximum value for each field is set to 15 and the time is coarse grain, there are approximately 4 million $(10^5 * 5)$ possible observations that can be made. As a comparison: In the work of Farrahi [] only 512 $(4^3 * 8)$ different observations are possible and 2856 days of data for 68 people is available.
In table 4.1 an overview is given how many unique words are actual observed for the different houses and how many words are totally observed $(48 * \#$ of days). One can see that the number



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

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.

HouseNr	1	2	3	4	5
# of days	142	98	89	63	73
words	6816	4704	4272	3024	3504
unique words	2147	1644	1764	1068	1087

Table 4.1

of unique observations scales with the number of days available for each house. This shows that there are a lot of unseen observations in the data.

The given feature representation, with the two variation of the time dimensions, fine grain and coarse grain, are used in the following chapters. There are much more possibilities to describe the feature. An overview of different possibilities is given in the Future work (chapter 7).

Chapter 5

Topic Models

In the first section of this chapter the general idea of topic models is introduced. This section is followed by a description of how this kind of models can be applied the sensor data. After that the extension of the LDA model 'LDA-Gaussian' is given, including the EM-algorithm that is used to determine the model parameters. This chapter is finalized with the description of the LDA-Poisson model.

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 'Finance' 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 it is described how this model can be used with the sensor data, that is gained in the different houses.

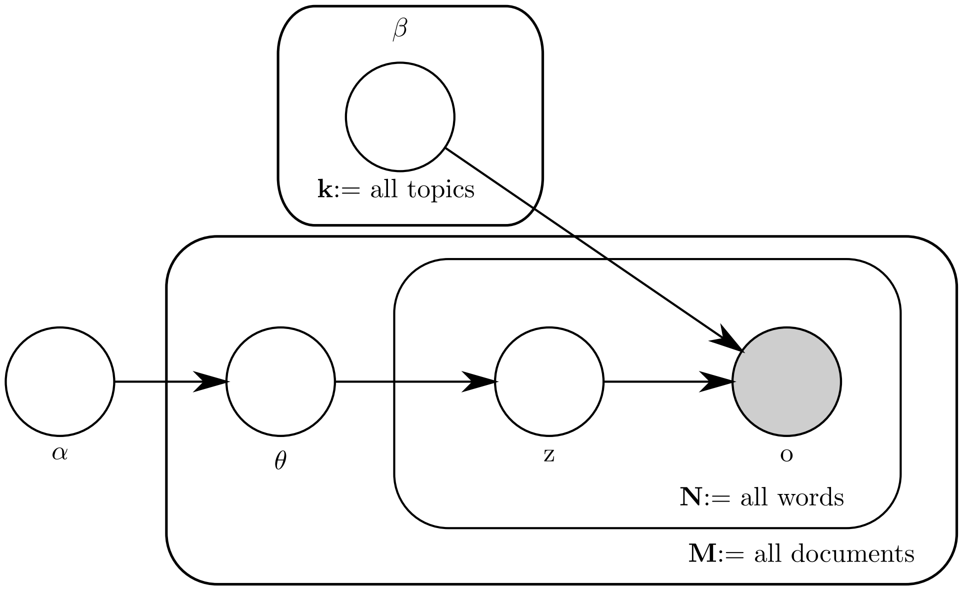


Figure 5.1: Graphical representation of the LDA model

5.2 Topic models with Sensor Data

In order to employ the topic model with the sensor data, first the different levels of description are introduced and relate to the terms of document classification.

- **Dataset/Corpus**: One dataset C describes the sensor data that is obtained 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, which is a natural choice because daily routines are a good indication for behavior patterns. A day can be compared with a document in a Corpus.
- **Observations/Words**: Finally every day is build of a set of observations, which are already introduced in chapter 4. Observations can be roughly compared with words in document classification. A different name for this observations is 'Artificial words'.

There are some differences between the data that is used for topic detection in text documents and in data obtained from binary sensors.
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 the case of sensor data, two observation that are similar to each other, are more likely to refer to the same topic. So for example if the topic "preparing breakfast" has the observation $o_n = 2, 4, 0, 3, 2, 2$ a similar observation like $o_n = 2, 4, 0, 4, 2, 2$, where only the fourth value of the vector is changed, may also refer to the same topic. With a BOW model, where every unique observation forms a new dimension, this similarity cannot directly be captured. The only way to capture this relationship is by observing the similar observations with another observation. Say observation A and B are similar to each other. So if observation A is seen with observation C and also B is seen with C, then it might be possible that the LDA algorithm assigns the same topic for A and B. A lot of data is necessary to find these kind of relations.
Another difference is that in the topic model LDA, it is assumed that the order of the words in the text is not important. However with sensor data the time when an observation is made is of big influence. By adding the time value to the observation we can overcome this problem. The fine grain time representation (see chapter 4) will contain more information about the sequential

order of the observations than the coarse grain representation.

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 an example is shown how a topic model can assign topics to the different observations. It is not necessary that two observations lay close to each other if they are assigned to the same topics. So the spatial relationship between the observation is not taken into account. Observations can be assigned to multiple topics.

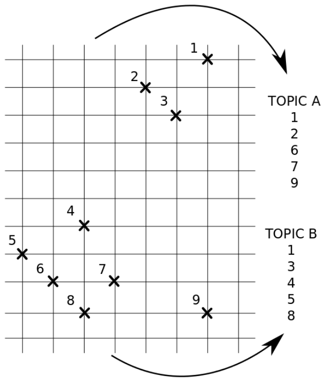


Figure 5.2: The Bag-of-words model with topic assignment.

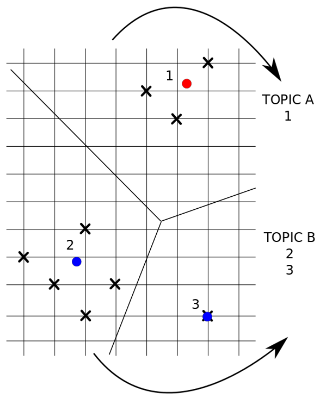


Figure 5.3: Feature space with k-means and topic assignment.

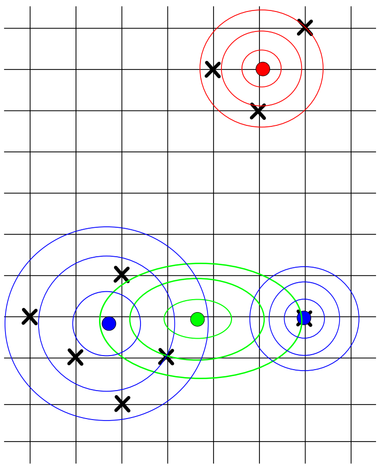


Figure 5.4: Gaussian Mixture Model with LDA. Topic A is marked with red. Topic B is marked with blue (clustering on forehand with GMM) and with green (combination of GMM and LDA)

To make sure that similar observations will be grouped together and appear in the same topic, the k-means algorithm can be applied to find clusters in the data. The id's of the centroids of the clusters then function as the 'artificial words'. The size of the dictionary of the BOW model is in this way reduced. Applying LDA on this reduced representation can lead to a better topic description. 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. Choosing the correct amount of clusters for the k-means algorithm can be difficult, as

it is also pointed out in the work of Casale [2]. And still it can occur that the clusters do not give a good representation of the data. All clusters have the same influence on the LDA model, which is not always desirable. A Gaussian Mixture model (GMM) can give a better description of the clusters and give a degradation of the topics. Every Gaussian distribution can be assigned to a topic with the LDA model. This is shown in figure 5.4 with the red and blue Gaussian distributions. The red distribution is assigned to one topic and the two blue distributions are assigned to another topic.
Combining the clustering and the Topic model directly together can lead to a more global representation of the topics. This is indicated with the green Gaussian distribution in the figure, which combines the two blue distribution into one.

5.3 LDA-Gaussian Model

In this section the LDA-Gaussian model is described. First the model is explained, which is followed by the description of the variational inference. This inferense step is necessary to calculate the model parameters. In this subsection also the two steps of the EM-algorithm are explained.

5.3.1 Model Description

For the model it is assumed that every day in a dataset can be represented as random mixture of latent topics, where every topic can be described as a distribution over observations. The following generative process is assumed 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 $i = z_n$.

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 $\mathcal{N}(\mu_{di}, \sigma_{di})$. With the feature representation given in chapter 4 the size of the dimension is fixed with $D = 6$. α represents the Dirichlet parameter and is vector of length k.

In figure 5.5 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, the optimal parameters need to be found. Therefore the probability for the dataset C needs to be maximized, with respect to the parameters α, μ and σ. This probability looks like this

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

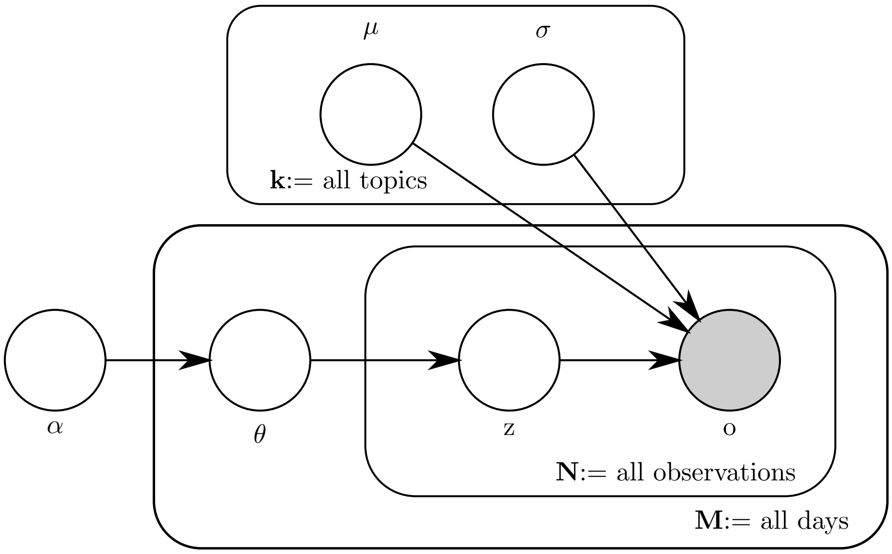


Figure 5.5: Graphical representation of the LDA-Gaussian model

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 obtained 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 it is described how the optimal parameters for the model given a data set can be found.

5.3.2 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 a convexity-based variational algorithm is applied, so that the log-likelihood of a given dataset can be approximated. 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 probability $p(z_i|o_n)$ and is given as a $k \times N$-matrix for every day m. The graphical representation of the model is shown in figure 5.6.

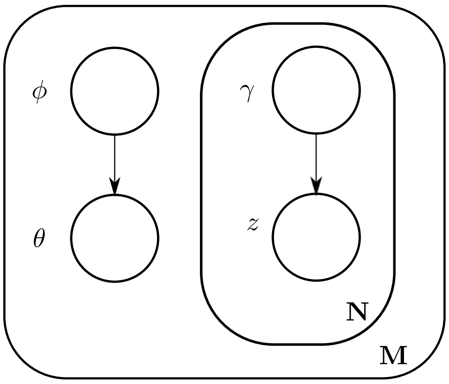


Figure 5.6: Approximation of the model.

Given the variational distribution for the approximate model the lower bound of the log-likelihood can be estimated with the Jensen inequality

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

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

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

An EM-process is 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.

Now a more detailed description on both of these steps is given:

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

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

From this equation the derivative can be taken and set it to zero. This leads to the first update function

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

A similar approach is handled for γ. Again all terms of equation 5.7 that contain this variable is selected and the derivative is set 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 again all terms of the likelihood that contains the variable α are selected. 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 the Newton-Rhapson method can be used to calculate the optimal α.

5.4 LDA-Poisson Model

The feature description in chapter 4 generates positive integers. Small values are more likely to occur in the observations and that is why a Poisson distribution is a better choice for the description of the topics. In figure 5.7 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

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

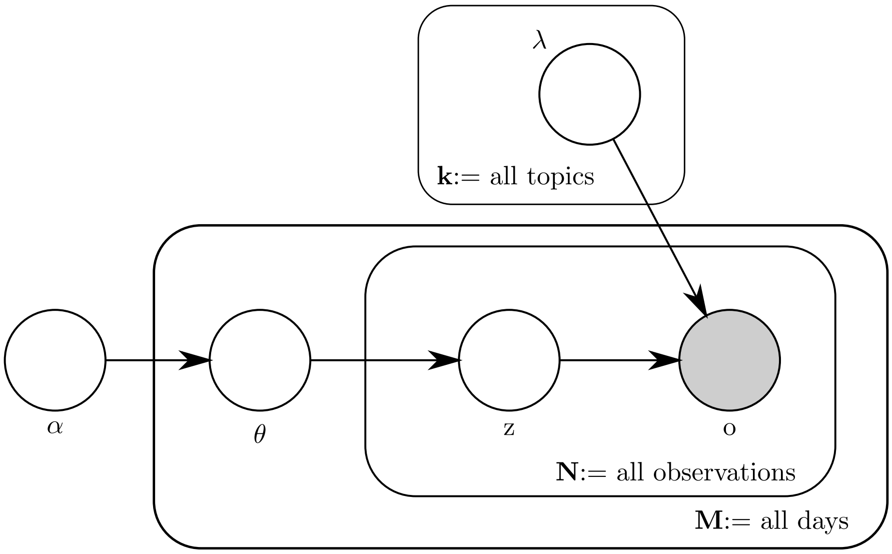


Figure 5.7: Graphical representation of the LDA-Poisson model.

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.

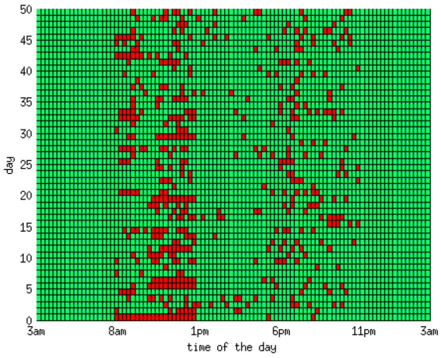


Figure 6.1: Topic distribution for 50 days for the Bag-of-Words model

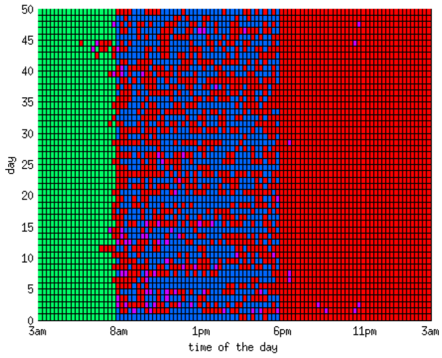


Figure 6.2: Topic distribution for 50 days for the k-means model

Figure 6.3

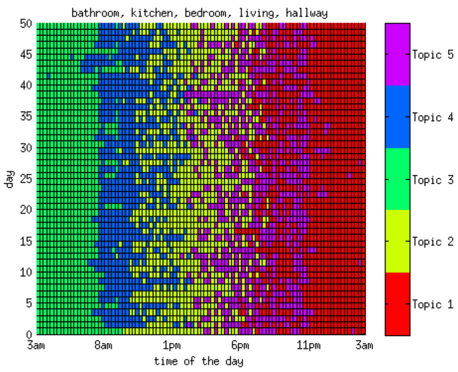


Figure 6.4: Topic distribution for 50 days

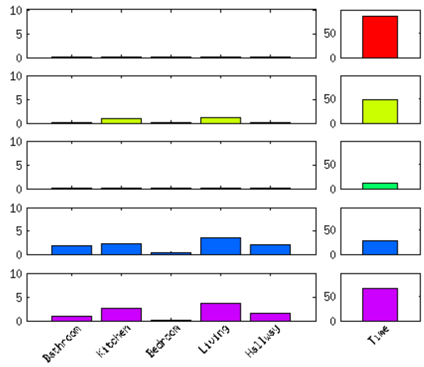


Figure 6.5: Visualization of the topics

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

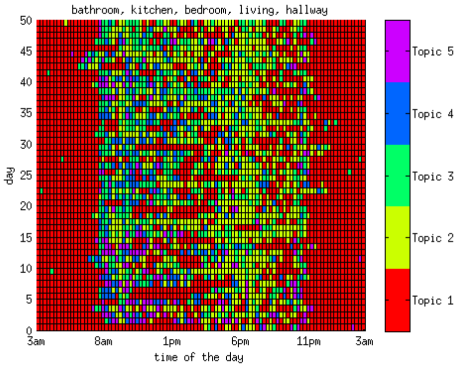


Figure 6.7: Topic distribution for 10 days

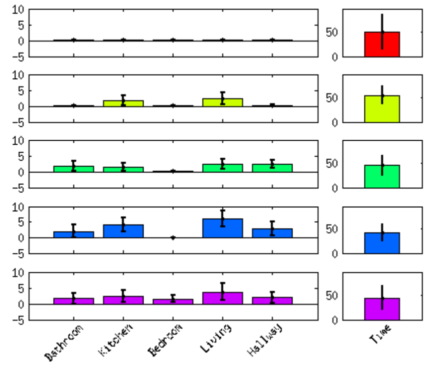


Figure 6.8: Visualization of the topics

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

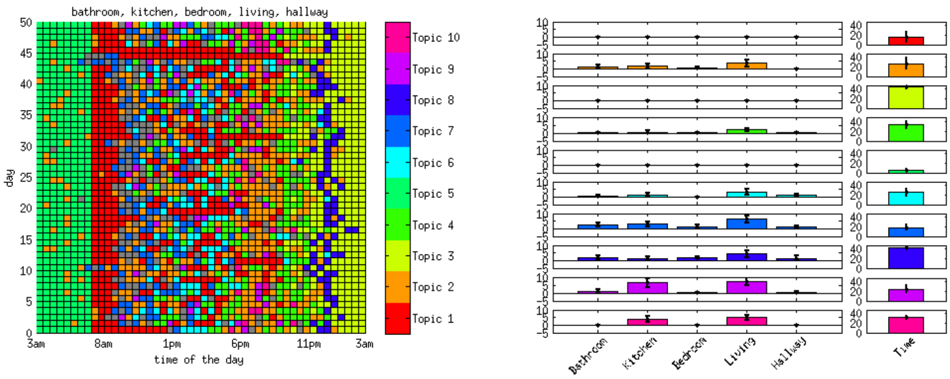


Figure 6.10: Topic distribution per day and the Topic visualization for LDA-Gaussian. HouseNr=1

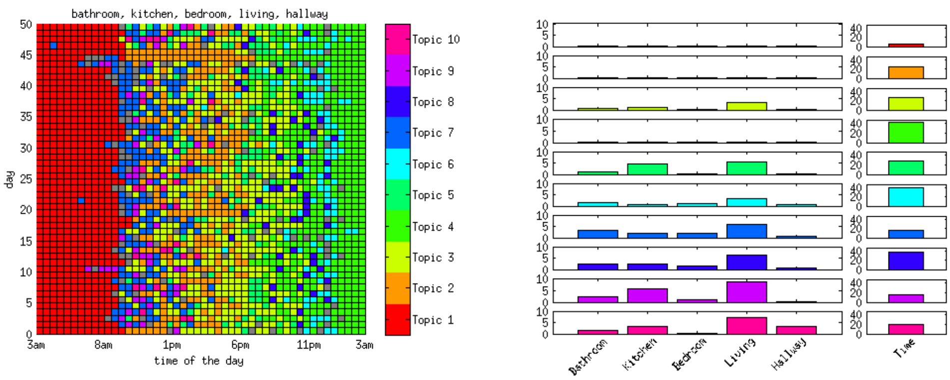


Figure 6.11: Topic distribution per day and the Topic visualization for LDA-Poisson. HouseNr=1

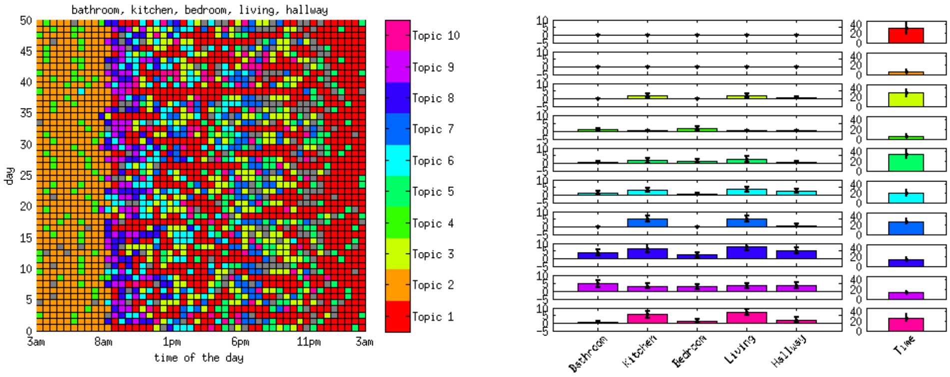


Figure 6.12: Topic distribution per day and the Topic visualization for LDA-Gaussian. HouseNr=2

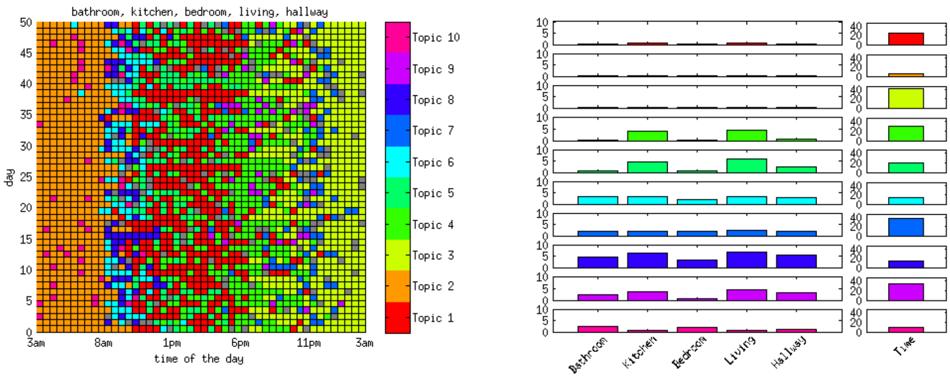


Figure 6.13: Topic distribution per day and the Topic visualization for LDA-Poisson. HouseNr=2

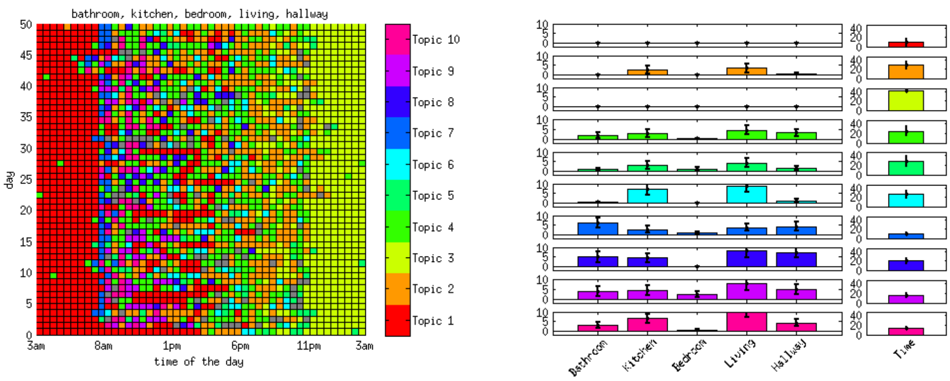


Figure 6.14: Topic distribution per day and the Topic visualization for LDA-Gaussian. HouseNr=3

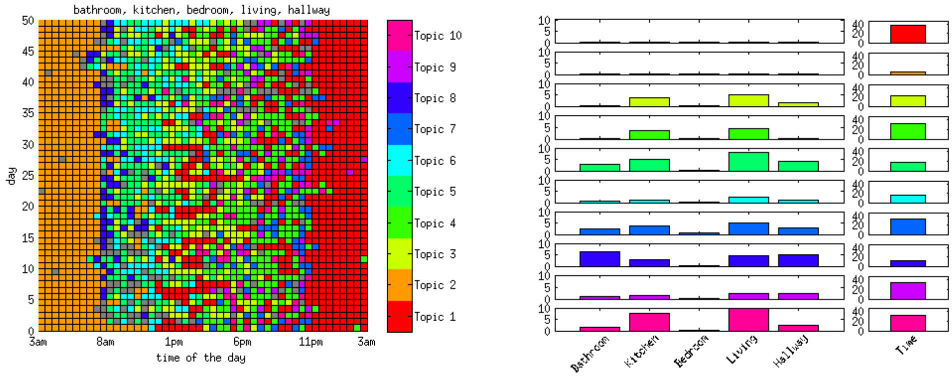


Figure 6.15: Topic distribution per day and the Topic visualization for LDA-Poisson. HouseNr=3

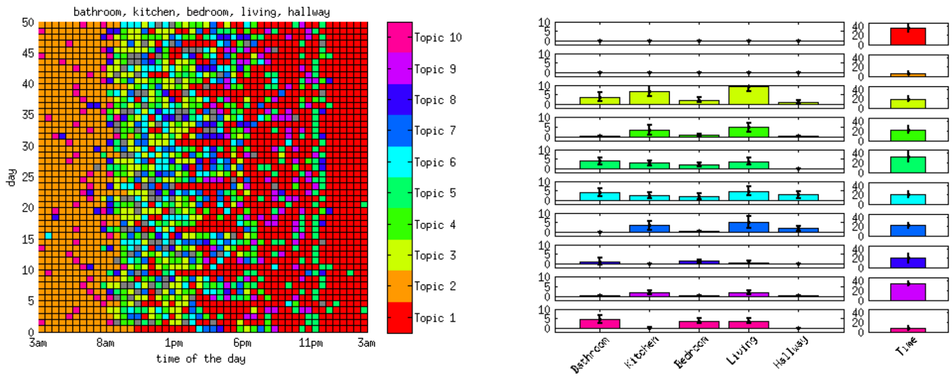


Figure 6.16: Topic distribution per day and the Topic visualization for LDA-Gaussian. HouseNr=4

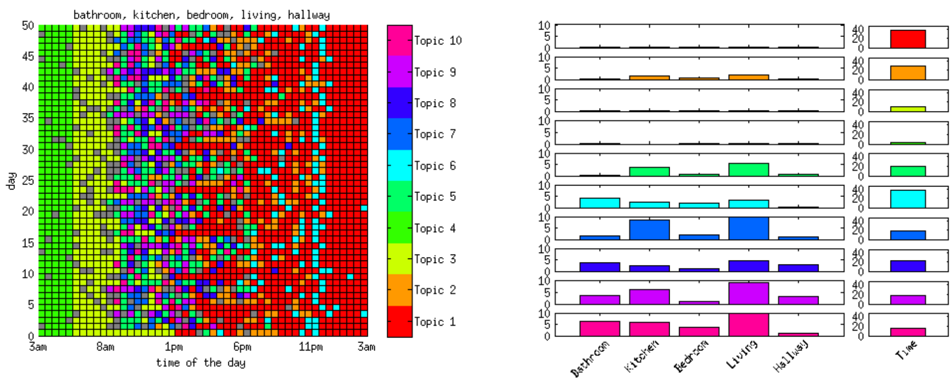


Figure 6.17: Topic distribution per day and the Topic visualization for LDA-Poisson. HouseNr=4

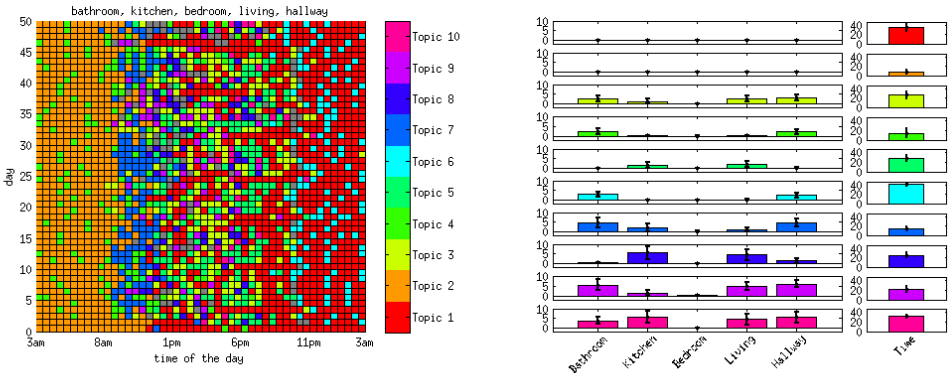


Figure 6.18: Topic distribution per day and the Topic visualization for LDA-Gaussian. HouseNr=5

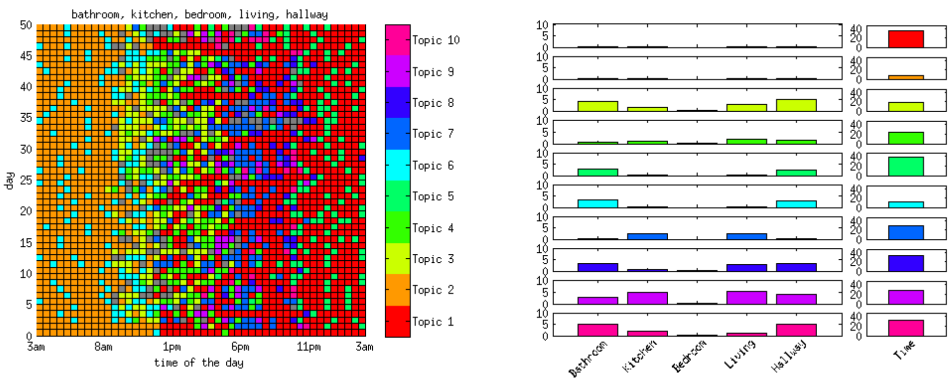


Figure 6.19: Topic distribution per day and the Topic visualization for LDA-Poisson. HouseNr=5

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.

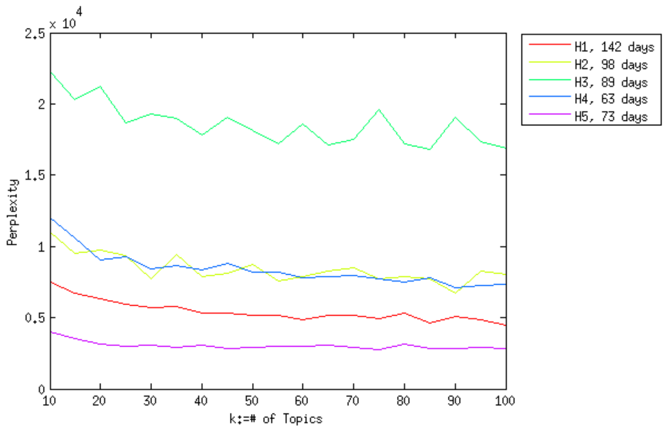
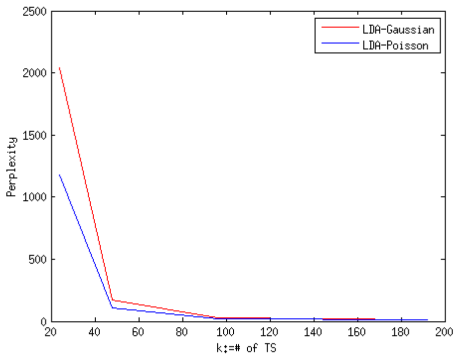


Figure 6.20: Perplexity for different initializations of Topics for 5 different House

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.



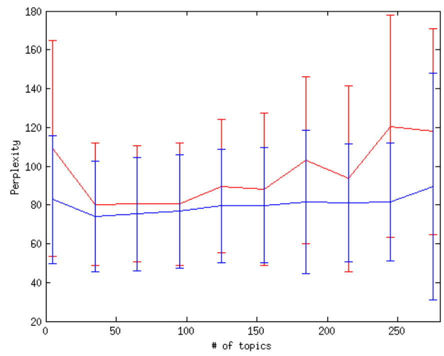


Figure 6.21: Perplexity for LDA-Gaussian and LDA-Poisson with different amount of time-slices

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

Figure 6.23

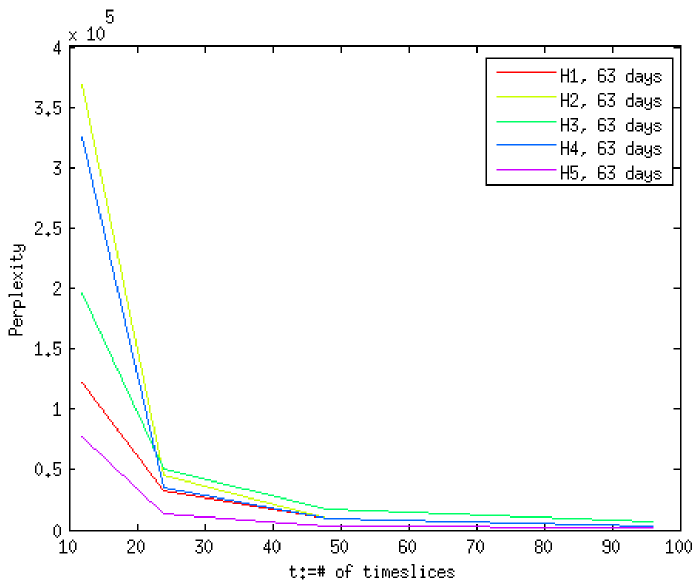


Figure 6.24: Perplexity (x-axis) for the hold-out-set for different amount of time-slices (y-axis)

Chapter 7

Future Work

Chapter 8

Conclusions

Bibliography

[1] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March 2003. 1, 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. 2, 14

[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. 2, 4

[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. 1, 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. 1, 2, 4, 9

[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¿. 1, 3

[8] Tâm Huynh, Mario Fritz, and Bernt Schiele. Discovery of activity patterns using topic models. In *Proceedings of the 10th international conference on Ubiquitous computing*, UbiComp '08, pages 10–19, New York, NY, USA, 2008. ACM. 2

[9] 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. 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] Dinh Q. Phung, Brett Adams, Kha Tran, Svetha Venkatesh, and Mohan Kumar. In *PerCom*, pages 1–9. IEEE Computer Society. 4

[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. 1

[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. 1, 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. 1, 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