Building an Occupancy Model from Sensor Networks in Office Environments

Federico Castanedo, Diego López-de-Ipiña
DeustoTech - University of Deusto
Avda. Universidades 24. Bilbao, Spain
{fcastanedo,dipina}@deusto.es

Hamid Aghajan
Dept. of Electrical Engineering
Stanford University
aghajan@stanford.edu

Richard Kleihorst
Vito NV and Ghent University
IBBT Belgium
richard.kleihorst@vito.be

***Abstract*—The work presented here aims to answer this question: Using just binary occupancy sensors is it possible to build a behaviour occupancy model over long-term logged data? Sensor measurements are grouped to form artificial words (activities) and documents (set of activities). The goal is to infer the latent topics which are assumed to be the common routines from the observed data. An unsupervised probabilistic model, namely the Latent Dirichlet Allocation (LDA), is applied to automatically discover the latent topics (routines) in the data. Experimental results using real logged data of 24 weeks from an office building, with different number of topics, are shown. The results show the power of the LDA model in extracting relevant patterns from sensor network data.**

I. Introduction

Mining the huge amount of data obtained from a sensor network in a modern office building presents a significant challenge for pattern recognition and behaviour analysis. The results of this analysis could provide several benefits in different ambient intelligence areas and domains. In this paper, we focus on the long-term analysis of people behaviour's in an office environment, meaning by long-term a time-line bigger than six months. This temporal constraint relies on the assumption that several interesting patterns could only be detected when behaviour data is analyzed over a long time. Thus, for finding patterns it is necessary to employ techniques and algorithms which are able to cope with the huge amount of data provided by the sensors, in a scalable manner, and following an unsupervised way.

The work presented here employs the Latent Dirichlet Allocation (LDA) [1] model in order to find interesting patterns from the data obtained from the environment. LDA is a generative and unsupervised model for collections of discrete data and it is also known as a probabilistic topic model. LDA is different from other clustering techniques in that it assigns multiple topics to each document. Viewing the activity history inside the building as documents with each word corresponding to some part of the activity makes it possible to formulate the problem of discovering the set of topics that are used in a collection of documents (corpus). The idea behind this comparison is that the set of topics would correspond to the set of most common patterns in the environment being monitored (documents) which are drawn from the observed activities (words).

One of the objectives and aims of ambient intelligence applications is to understand what is going on in an environment, which could be derived from sensory observations, and respond appropriately to the identified solution. In LDA the semantic properties of activities (words) and activity history (documents) are expressed in terms of probabilistic routines (known as topics) which are learned following an unsupervised scheme. Therefore, semantic relationships between words-topics and topics-documents are interpreted in terms of probability distributions.

The kind of sensors employed in ambient intelligence applications usually provides a huge amount of data making necessary to employ some filtering or dimensionality reduction techniques. LDA could also be seen as a dimensionality reduction technique since it transforms a document into a set of topics and each topic is drawn from a multinomial distribution of words. Thus it is possible to describe each document from the most likely topics and each topic with the K most probable words. Also, the complexity of the LDA model, once it has been trained, is linear with the number of topics learned and the size of the corpus (collections of documents), making it suitable to process a large amount of data.

The application of the LDA model in this scenario allows to infer in a *generative way* a set of routines (topics) on which depends a set of daily patterns of activities (documents). We also argue, that it could be possible to extend or reduce the word length and the document size in order to infer activities over different time intervals.

The work presented here aims to answer this question: Using just binary occupancy sensors is it possible to build a behaviour occupancy model over long-term logged data?

This paper continues as follows. In the next section, some related works are described and the differences with this work are highlighted. Section III describes the process of building the model from the obtained raw measurements. The results and the experiments carried out are shown in section IV. Finally, section V provides future research directions and concludes the paper.

II. Related works

Since Blei's LDA original paper [1], the model has been applied to different domains but mostly in the document processing domain using text collections. In [2] the authors use the LDA algorithm to analyze abstracts from PNAS and automatically extract the topics. In contrast to the original LDA model which uses variational inference they employed a Markov Chain Monte Carlo (MCMC) inference model. A good comparison between the different inference algorithms that could be employed in the LDA model is provided in [3].

In the computer vision domain, the LDA model has been applied to automatically infer the images' categories and understand the content of the images such as in [4] [5]. It has been also employed for learning human action categories in videos [6].

Recently, Farrahi and Gatica-Perez [7] [8] use LDA and Author-Topic Model [9] to discover daily location-driven routines from a massive dataset of mobile phones user's location. In the work presented here, a similar approach is employed to build the *bag-of-words* of the documents. However, their experiments focus on data which have different locations from each user, instead of this work which models the occupancy at one location.

Related to the office building domain, the work of [10] provides statistical sampling of occupancy features from a very large dataset (MERL dataset [11]) with no ground-truth or outside knowledge. The MERL dataset [11] is composed of people movements at the MERL building for a period of a year (from March 21 2006 to July 2 2008) and recorded over a sensor network of 200 wireless motion sensors. The full dataset contains over 30 million raw motion sensor data logs from two floors. Their approach is based on information theoretic measures and graph-cuts and their goal was to identify potentially important events within the organization.

A review of temporal pattern mining algorithms used to discover patterns in sensor networks is provided in [12]. They proposed a modified T-Pattern algorithm [13] and tested it using the MERL dataset. On the smart-home environment domain, the work of [14] introduces temporal information in the theory of evidence to perform the activity recognition problem.

The work presented here, follows a novel approach, since the sources of information are sensor measurements which are grouped to form artificial words (activities) and documents (a set of activities over time) which generate a corpus representing the dynamic bahaviour of the monitored environment. This paper, extends the work presented in [15] using a different word length and more experiments.

III. Building the LDA model

In a generative model the goal is to find the best set of latent variables (in this case, topics or routines) that can explain the observed data, assuming that the model actually generates the observed data. The LDA model is based on the *bag-of-words* assumption, that is, the only information relevant to the model is the frequency of words. Thus, in Bei's original model, the order of the words in each document is not taken into account. To easily overcome this assumption, in this work, the obtained words are segmented in 9 different time intervals according to the time of the day which generates the activity. This produces a bigger vocabulary, however the *bag-of-words* assumption persists inside the words generated from each time interval. But, it makes sense since the granularity of the activities we focus on is based on routines instead of low-level activities. So, the input to the model is the *bag-of-words* representation of a collection of text documents, where documents D are represented as a sparse vector of $|W|$ non-negative counts from the words of a vocabulary V.

Given D documents that could be expressed with T topics over W words, it could be possible to represent $P(W|T)$ with a set of T multinomial distributions ϕ_t over the W words. $P(W|T)$ represents the probability distribution over words W given a topic T. LDA models each document d as a mixture θ_d over T latent topics, where each topic ϕ_t is a multinomial distribution over the $|W|$ word vocabulary. LDA assumes a Dirichlet prior distribution on the parameters θ_d and ϕ_t.

θ_d is a $|D|$xT matrix of document-specific mixture weights for the T topics each of them drawn from a Dirichlet(α) prior distribution with hyper-parameter α. The parameter α provides information about how semantically diverse documents are (in this case daily occupancy patterns) with lower values indicating a higher diversity.

ϕ_t is a $|W|$xT matrix of word-specific mixture weights from the words W of the vocabulary V for the T topics, each of them drawn from a Dirichlet(β) prior distribution with hyper-parameter β. The parameter β provides information about how similar the different topics are, it provides the log-likelihood of $prob(word|topic)$ for each one of the T topics.

Given the parameters α and β, the joint distribution of a topic mixture θ, a set of T topics z, and a set of N words w is given by:

$$p(\theta, z, w|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^{N} p(Z_n|\theta) p(w_n|Z_n, \beta) \qquad (1)$$

the computation of equation (1) parameters is intractable in general, due to the coupling between θ and β and different approximate inference algorithms, such as Gibbs sampling, Laplace approximation, variational approximation and Markov Chain Monte Carlo [16] should be employed. For a good comparison between the different inference algorithms that could be used in LDA, see [17]. In this work a variational $Expectation - Maximization$ inference algorithm, as detailed in [1], is used in order to find the parameters θ_d, α, ϕ_t and β.

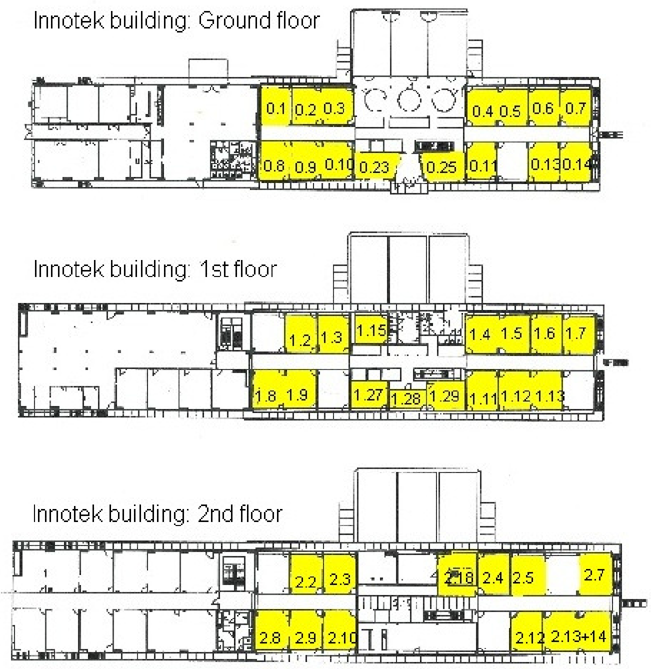

Fig. 1. Plan of the three floor Innotek building. Rooms highlighted in yellow are those which have some activity information.

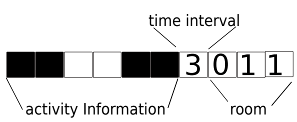

Fig. 2. Example of word (IIOOII3011) which indicates an activity pattern of someone in the room for 10 minutes, out for 10 minutes and in again for other 10 minutes at time interval 3 (between 7:00 and 9:00 am.) in the room 0.11

When the LDA model is trained it could be used to infer the topic distribution of new documents given the document-topic Dirichlet prior and the words distribution of each topic given the topic-word learned distribution. The learned distributions could be used for classifying new documents or measuring similarities between documents.

The information of the words W which formed the documents D, the employed number of topics T and the probabilities of some learned distributions is shown in the next section.

IV. Experiments

One of the main questions that arise when dealing with the LDA model is what is a word W and what set of words are used to make up the documents D. This section reports the obtained results with the Innotek dataset using different modeling schemes.

The data for the Innotek dataset were obtained from monitoring the Innotek building in Geel, Belgium (see Figure 1) from March 12 2010 to August 28 2010.

The occupancy data is captured by centrally mounted high-quality PIR-based sensors that are part of the Philips system to install the automatic lighting. Detection of motion will trigger the lights which will remain switched on for at least 10 minutes. If no motion has been detected in this period, the lights will be turn-off again. Innotek has arranged for connecting these sensors to the bus of their Johnson Controls climate management system, which enables their use to also control the room temperature. The climate management system will set the room to *hibernation mode* if no occupancy has been detected for a certain time interval. Once occupancy is detected, the room climate will go to the comfort level as set locally per room by the inhabitants.

As the Johnson Controls sensor bus can be read out at a central point we have been able to log the data of most rooms of the building on a 1 minute accurate timescale. The events are based on integration of detections over the 1 minute interval and indicate the occupancy with a high level of confidence.

The raw obtained measurements are mapped into *words* in order to generate activity information. Each of the 30-minutes duration words (see Figure 2) that could formed the documents is composed of 10 digits:

- 6 slots of 5 minutes of duration. Each slot means if that room is occupied or empty during that interval, it is taken the maximum between each 5 minutes occupancy information. Please, note that these 5-minutes slots do not overlap between each other, providing sequential information over time. (6 digit)
- a number between 1 and 9 indicating the time interval: from 0:00 to 6:00 (1), 6:00 to 7:00 (2), 7:00 to 9:00 (3), 9:00 to 11:00 (4), 11:00 to 12:00 (5), 12:00 to 17:00 (6), 17:00 to 19:00 (7), 19:00 to 21:00 (8) and from 21:00 to 24:00 (9). (1 digit)
- the room number (3 digits).

A. How many Routines (Topics)?

To establish the best number of topics T is one of the main tasks which must be taken into account in the LDA model. Too many topics may produce random results of words which may be difficult to interpret. On other hand, too few topics usually result in a very broad groups. A common approach is to use the number of topics that leads to best generalization performance over unseen new documents (validation data), known as *perplexity*. This measure is used by convention in language modeling and the standard one is defined as the reciprocal geometric mean of the likelihood of a test corpus given a model. However, there are different ways to calculate the perplexity, we refer the reader to [18] for more information.

Formally for M documents, standard *perplexity* measurement is defined as:

$$perp = exp(-\frac{\sum_{m=1}^{M} \log p(W_m|Lda)}{\sum_{m=1}^{M} N_m}) \qquad (2)$$

where $p(W_m|Lda)$ is the probability of the unseen set of words in document m given the LDA trained model and N_m the number of words in each document.

The optimum number of latent topics also depends on the underlying nature of the data, that is, as in other clustering algorithms, the optimum number is related with the clusters that the data should have.

The approach followed next in the experiments is to find the value of T which provides a lower perplexity indicator and then obtain the K most likely words that composed the most relevant topics.

B. Intra-Room Occupancy Modeling

First, the LDA model is trained for each room, using a corpus formed by the documents having occupancy information of each room for each day. From all the rooms of the Innotek building, just those which provide occupancy information were employed, these rooms are highlighted in Figure 1. Therefore a document corresponds to the set of activities (words) that took place in one room over one day.

In these experiments the vocabulary size of each room's model is formed by 2^6 (patterns of occupancy) times the time interval (9), that is $64 * 9 = 576$ different words. The empty words at interval 1 and 9, that is at night or early in the morning were removed and not used for training the model, since they are usually considered as stop-words. Only information about weekdays was taken into account, giving a total amount of 24 weeks. The occupancy data were split between 12 different subsets of training and validation. A training set of 22 weeks (110 documents) were employed in order to train the LDA model. The validation set of 2 weeks (10 documents) performs the inference using the model over the unseen data. Perplexity measurements were obtained from 12 different subsets using T values of 7,8,9,10,15,20,25,50,100 and 150. The graphs of how the number of topics T affects the perplexity values of rooms 0.11, 1.2, 2.7 and 2.9 are shown in Figure 3

It seems that the obtained results are consistent with the previous classification of the time intervals and the model obtains the clusters accordingly. As an example, Table I shows the 3 most probable words for each obtained topic of room 0.11 (which corresponds to number 10) with $T = 10$. A picture of the frequencies of the room 0.11 is also shown in Figure 5.

C. Inter-Room Occupancy Modeling

Second, the LDA model was trained using the information of all the rooms from each floor with the purpose to obtain related patterns between rooms. In this scheme each document is composed of the daily information from all the rooms of one floor. For instance, for floor 0, 15 different rooms were employed: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.10, 0.11, 1.12, 0.13, 0.14, 0.23 and 0.25 to train the model. The vocabulary size of this scheme is formed by 23616 words and the number of documents of the training set is ≈ 1450 since not all the rooms provide information over the whole data, which in that case would be 1500 documents. The obtained perplexity values from each number of topics using the 12 cross validation sets for the floor 0 are shown in Figure 4.

From the obtained values shown in Figure 4 it seems that a value of $T = 50$ provides a good choice, so the 3 most probable words of some of the obtained topics with $T = 50$ are shown in Table II. The activity of room 24 (which corresponds to room 0.25 in the map) provides an interesting occupancy behaviour (see Topic 5,6,8,13 and 34 of Table II). For instance, topic 5 indicates that there are in and out activities within a slot of 15 minutes at the time interval 2 (between 6:00 and 7:00 am) with a relative high probability. Moreover topics 6, 8, 13, and 24 also indicate that in this room there are a lot of interesting activity patterns at intervals 3, 5 and 7. On the other hand, topics 35, 36, 40 and 3 indicate a similar occupancy pattern between rooms 13, 12, 6 and 1 (rooms 0.14, 0.13, 0.7, and 0.2 on Figure 1), that is without occupancy between 11:00 and 17:00.

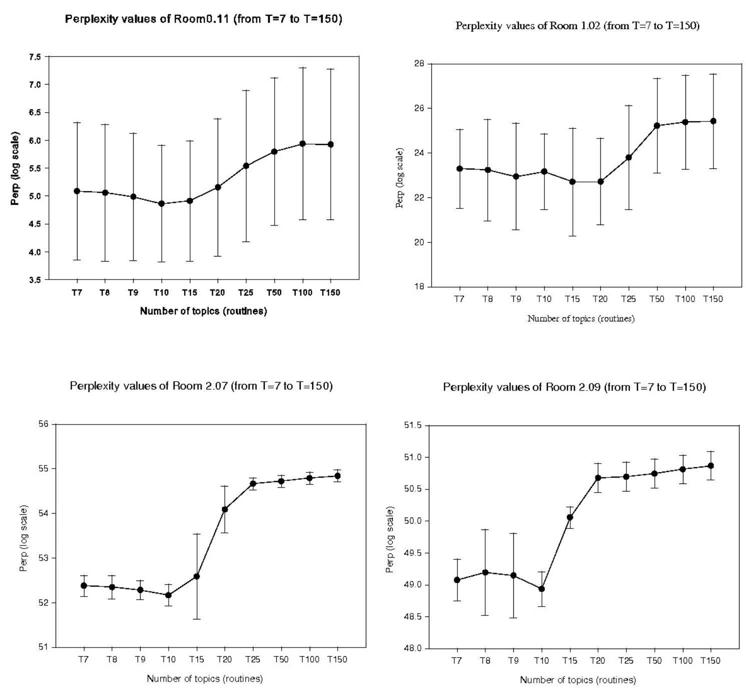

Fig. 3. These graphs show the average and standard deviation (log scale) of the 12 cross-validation executions of the obtained perplexity values from the LDA model trained using the information of the following rooms: (a) Room 0.11, floor 0, it is shown that a topic value of $T = 10$ provides the best results. (b) Room 1.02, floor 1, it is shown that a topic value of $T = 10$ provides results with less dispersion than T=9 and T=15. (c) Room 2.07, floor 2, it is shown that a topic value $T = 10$ provides the best results. (d) Room 2.09, floor 2, it is shown that a topic value $T = 10$ provides the best results.

TABLE I

3 MOST PROBABLE WORDS FOR EACH OBTAINED TOPIC OF ROOM 0.11 WITH $T = 10$

Topic 1		Topic 2		Topic 3		Topic 4		Topic 5	
word	p(w|topic)	word	p(w|topic)	word	p(w|topic)	word	p(w|topic)	word	p(w|topic)
0000006010	0.227865	0000007010	0.182136	IIIIII5010	0.202099	0000004010	0.154278	IIIIII6010	0.232004
0000008010	0.139317	0000008010	0.145584	IIIIII6010	0.198097	IIIIII6010	0.134526	0000008010	0.120875
IIII003010	0.003410	0IIIII6010	0.018911	0000008010	0.128221	0000008010	0.124059	IIIIII7010	0.056416
Topic 6		Topic 7		Topic 8		Topic 9		Topic 10	
word	p(w|topic)	word	p(w|topic)	word	p(w|topic)	word	p(w|topic)	word	p(w|topic)
0000004010	0.222017	0000005010	0.172389	0000006010	0.223353	IIIIII4010	0.242440	0000006010	0.292585
0000007010	0.164146	0000007010	0.138207	0000007010	0.120996	0000005010	0.131204	0000008010	0.137381
IIIIII3010	0.059309	0000008010	0.130942	0000008010	0.116019	0000008010	0.130029	000III5010	0.016790

Fig. 5. Picture of occupancy frequencies for room 0.11 over the whole data (24 weeks) for the 7200 week minutes. Each cell corresponds to a 10 minute interval. Rows indicate the frequency of occupancy in gray-scale (the most likely occupancy is black) and columns the time evolution.

TABLE II

3 MOST PROBABLE WORDS FOR SOME OBTAINED TOPICS OF THE FLOOR 0 WITH $T = 50$

Topic 3		Topic 5		Topic 6		Topic 8		Topic 13	
word	p(w|topic)	word	p(w|topic)	word	p(w|topic)	word	p(w|topic)	word	p(w|topic)
0000005013	0.205769	000III2024	0.257175	00IIII3024	0.171187	IIIII07024	0.546621	II0III5024	0.233868
0000006013	0.203926	0III002024	0.128993	IIII003024	0.130289	III00I7024	0.136695	IIII0I5024	0.233568
0000008013	0.131598	0000II5024	0.128864	III0007024	0.076230	IIIIII5024	0.000555	0IIIII3024	0.076689
Topic 23		Topic 34		Topic 40		Topic 35		Topic 36	
word	p(w|topic)	word	p(w|topic)	word	p(w|topic)	word	p(w|topic)	word	p(w|topic)
IIIIII5002	0.203315	00IIII5024	0.177317	0000005006	0.192050	0000005012	0.210557	0000005001	0.232236
IIIIII6002	0.202301	III0005024	0.177817	0000006006	0.181591	0000006012	0.216557	0000006001	0.212460
0000008002	0.139457	IIII0I3024	0.088922	0000008006	0.131467	0000007012	0.131598	0000008001	0.130942

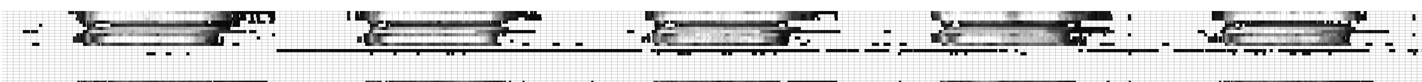

Fig. 6. Picture of occupancy frequencies for the floor 0 over the whole data (24 weeks) for the 7200 week minutes. Each cell corresponds to a 10 minute interval. Rows indicate the frequency of occupancy in gray-scale (the most likely occupancy is black) and columns the time evolution. Each line correspond to the occupancy information of one room.

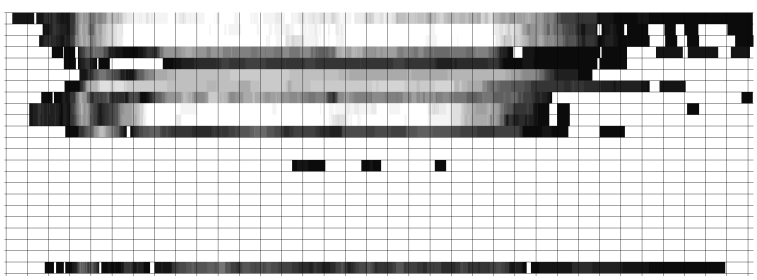

Fig. 7. Zooming in the occupancy frequencies for the floor 0 on Monday over whole data (24 weeks). Each cell corresponds to a 10 minute interval. Rows indicate the frequency of occupancy in gray-scale (the most likely occupancy is black) and columns the time evolution. Each line correspond to the occupancy information of one room.

V. Conclusion and future work

The experiments carried out with the provided data employed a word length of 30 minutes which is composed by the occupancy information of 6 intervals by 5 minutes of duration. However different word lengths could be employed in order to discover routines from different activity granularity.

Despite the LDA model seems to be a good choice for building a long-term occupancy model of sensor's data, more efforts and experiments should be taken with the aim of establishing the best representation and parametrization of the obtained data. It seems reasonable to change the words length and the number of topics with the objective of detect routines with different granularities. A common agreement in the document modeling domain is that the number of topics T should be influenced also by

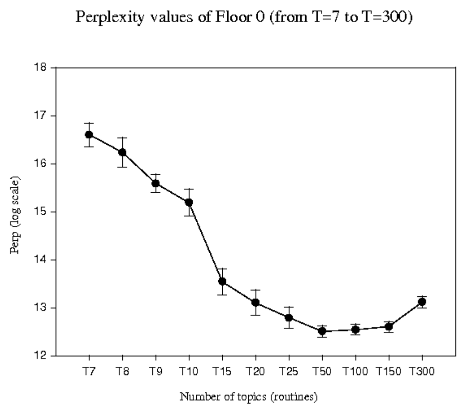

Fig. 4. Obtained perplexity values from the LDA model using all the rooms of floor 0. It is shown that T=50, T=100 and T=150 provide reasonable good results respect to others T values. The average and standard deviation of the 12 cross validation executions are shown (log scale).

the level of information that the application would like to provide. That is, over the same corpus some people could be interested in a broad range of topics, while others not. This could be an interesting idea to explore and research in the sensor network domain. In this work, for the intra-room occupancy modeling a value of $T = 10$ seems to provide the best perplexity (see Figure 3), which it is also consistent with the situation of having 9 different time intervals modeled a priori in the *words* construction phase. This value of $T = 10$ in the intra-room modeling indicates that the LDA trained model with 10 routines provides the best generalization over this test data. On the other hand, regarding floor 0, a number of $T = 50$, $T = 100$ and $T = 150$ routines provides good generalization results. In this case, the LDA model of floor 0 represents the routines of 15 different rooms (over 9 different time intervals), which also seems to be consistent with the obtained perplexity values of $T = 150$ (see Figure 4), since they provided a good result compared to other values of T.

One of the main advantages of the application of this unsupervised probabilistic model to the office environments could be the intelligent use of energy and a way to provide smarter and eco-friendly buildings.

Finally, there are several modifications to the original LDA model which could also be useful in this kind of environments, such as the dynamic topic models and may be of interested to complement this work.

Acknowledgments

The authors want to thank Erik Degroof and Luc Peeters of Innotek Belgium for their cooperation and the use of their dataset and David Ausin for his help with the experiments. The first and second authors are partially supported by the Spanish MICINN project TALIS+ENGINE (TIN2010-20510-C04-03).

References

[1] D. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," *The Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[2] T. Griffiths and M. Steyvers, "Finding scientific topics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. Suppl 1, p. 5228, 2004.

[3] A. Asuncion, M. Welling, P. Smyth, and Y. Teh, "On smoothing and inference for topic models," in *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. AUAI Press, 2009, pp. 27–34.

[4] J. Sivic, B. Russell, A. Efros, A. Zisserman, and W. Freeman, "Discovering object categories in image collections," *ICCV*, 2005.

[5] E. Horster, R. Lienhart, and M. Slaney, "Image retrieval on large-scale image databases," in *Proceedings of the 6th ACM international conference on Image and video retrieval*. ACM, 2007, pp. 17–24.

[6] J. Niebles, H. Wang, and L. Fei-Fei, "Unsupervised learning of human action categories using spatial-temporal words," *International Journal of Computer Vision*, vol. 79, no. 3, pp. 299–318, 2008.

[7] K. Farrahi and D. Gatica-Perez, "What did you do today?: discovering daily routines from large-scale mobile data," in *Proceeding of the 16th ACM international conference on Multimedia*. ACM, 2008, pp. 849–852.

[8] ——, "Discovering routines from large-scale human locations using probabilistic topic models," *ACM TIST*, vol. 2, no. 1, p. 3, 2011.

[9] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth, "The author-topic model for authors and documents," in *Proceedings of the 20th conference on Uncertainty in artificial intelligence*. AUAI Press, 2004, pp. 487–494.

[10] C. Connolly, J. Burns, and H. Bui, "Recovering social networks from massive track datasets," *IEEE workshop on Applications of Computer Vision, WACV*, 2008.

[11] C. R. Wren, Y. A. Ivanov, D. Leigh, and J. Westhues., "The MERL Motion Detector Dataset: 2007 Workshop on Massive Datasets," *MERL TR2007-069*, Mitsubishi Electric Research Laboratories, Cambridge, MA, USA. 2007.

[12] A. Salah, E. Pauwels, R. Tavenard, and T. Gevers, "T-Patterns Revisited: Mining for Temporal Patterns in Sensor Data," *Sensors*, vol. 10, no. 8, pp. 7496–7513, 2010.

[13] M. Magnusson, "Discovering hidden time patterns in behavior: T-patterns and their detection," *Behavior Research Methods*, vol. 32, no. 1, pp. 93–110, 2000.

[14] S. Mckeever, J. Ye, L. Coyle, C. Bleakley, and S. Dobson, "Activity recognition using temporal evidence theory," *Journal of Ambient Intelligence and Smart Environments*, vol. 2, no. 3, pp. 253–269, 2010.

[15] F. Castanedo, H. Aghajan, and R. Kleihorst, "Modeling and Discovering Occupancy Patterns in Sensor Networks using Latent Dirichlet Allocation," in *IWINAC*, 2011.

[16] M. Jordan, Z. Ghahramani, T. Jaakkola, and L. Saul, "An introduction to variational methods for graphical models," *Machine learning*, vol. 37, no. 2, pp. 183–233, 1999.

[17] A. Asuncion, M. Welling, P. Smyth, and Y. W. Teh, "On smoothing and inference for topic models," in *In Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence*, 2009.

[18] H. Wallach, I. Murray, R. Salakhutdinov, and D. Mimno, "Evaluation methods for topic models," in *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 1105–1112.