

A Dynamic Segmentation Based Activity Discovery through Topic Modelling.

Ihianle Isibor Kennedy*, Usman Naeem**, Abdel-Rahman Tawil[†]

School of Architecture and Computer Engineering, University of East London.
{u1051232*, u.naeem**, a.r.tawil[†]}@uel.ac.uk

Keywords: Probabilistic Latent Semantic Analysis (PLSA), Bag-of-words, Topic Model, Activity Discovery, Sensor Segments.

Abstract

Recent developments in ubiquitous and pervasive technologies have made it easier to capture activities through sensors. The “bag-of-words” topic models have been applied to discover latent topics in corpus of words. In this paper, we propose the Probabilistic Latent Semantic Analysis to discover activity routines. The framework we propose set latent topics as corresponding class labels and use the Expectation Maximization (EM) algorithm for posterior inference. The experimental results we present are based on the Kasteren dataset which validates our framework and shows that it is comparable to existing activity discovery approaches.

1 Introduction

Activity discovery from sensor sequences recently has provided great and viable means of support to the elderly and people with dementia. They are being used to provide assistance using Smart Home (SH) technologies by the recognition of the task being undertaken by the inhabitant [1]. Human activity recognition follows the process of data acquisition, segmentation, classification and recognition. In a typical SH system equipped with sensors, the activation and observation of sensors could mean that an activity has occurred. Therefore, the representation of the observed sensor sequences defines the activity. To classify these sensor observations to activity topics, we have adopted the Probabilistic Latent Semantic Analysis (PLSA) – a “bag-of-words” topic model. The “bag-of-words” PLSA topic model was developed with an aim of discovering words association with topic labels in a corpus of document in natural language and text processing [2]. Its assumption is that the words contained in a document have association with the hidden or latent topics. This mixture of the words relative to the topic is described by a probabilistic distribution. We conversely assume that activity labels would have probabilistic distribution relative to sensor observations. With this assumption relative to activity discovery, a sequence of sensor observations corresponds to the word segment in the documents and the underlying topics correspond to the activities.

Sensor observations typically are retrieved as a sequence of continuous data stream which when partitioned in segments of various sizes using temporal parameters yields a classification through which activity labels are discovered. Recent efforts have seen a selection of predefined number of topics or activity labels as obtained in parametric models [3] and the use of fixed sensor segment sizes of 60sec [4, 1] which otherwise restricts learning process and the discovery of new activities. Setting sensor segments to fixed and static sizes for activity classification process could affect activity transition and impact on the performance of the framework. Therefore, we propose through this paper a framework which would partition continuous sensor data into optimal segments and then discover the underlying activities through a classification process. This proposed framework, clusters sensor observations before the segmentation process. The contributions of this paper are of three folds. First, we present the dynamic and variable window sensor data segmentation. This segmentation model is designed to provide real time data acquisition and segmentation. Secondly, we propose “bag-of-words” Probabilistic Latent Semantic Analysis (PLSA) to discover activities over observed sensor data. Thirdly, we present experimental results which demonstrate the ability of the proposed framework to discover activity topics.

2 Related Work

Several attempts have been made in field of activity recognition to discover and recognise human activities like [5]. These efforts has seen researchers following probabilistic [6] logic [7] and Ontology [8] in the process of classification. The probabilistic Hidden Markov Model (HMM) is the most popular classification approach used. The HMM approach involves latent states generating observations through a stochastic process. Yamato et al [9] one of the earliest work used the HMM to recognise human actions from image sequences. Oliver et al [10] used Layered HMM framework to recognise office activities like conversations, making phone call and presentations. In [5] another extension of the HMM, the authors used Radio Frequency Identification (RFID) tags to recognise activities like doing laundry, preparing food, washing dishes etc. Another extension of the HMM is the Coupled HMM [11] used to recognise simultaneous human actions. Kasteren et al [4] used the HMM and the Conditional Random Field (CRF) to recognise leaving the house, Shower, Use Toilet etc. Whilst these models have recorded some positive achievements and popularity they also have some drawbacks. The reliance on

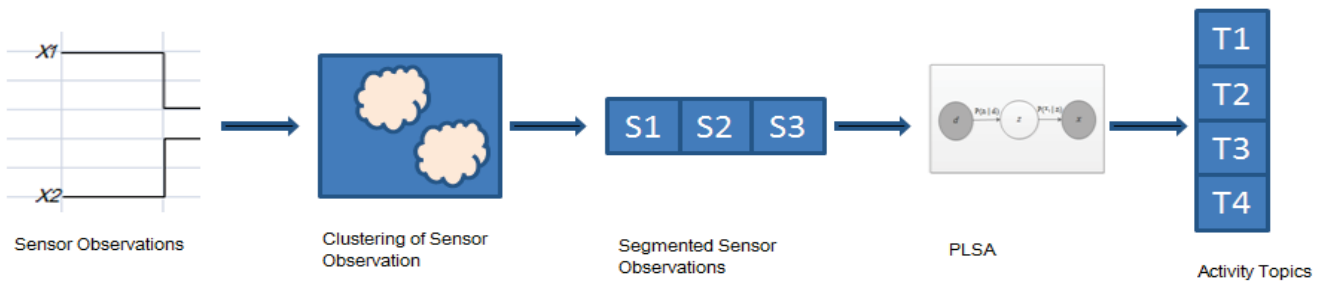


Fig 1. Sensor observations as inputs are clustered. They are partitioned into segments following their cluster composition. The PLSA through EM algorithm computes the posterior inference of the sensor segments as activity topics

the independence assumptions made by the 1st order HMM makes it unable to capture long range dependencies of the observations. Another popular approach in the human activity recognition is the machine learning discriminative technique Support Vector Machine (SVM). The approach relies on data points known as “support Vectors” for classification with reference to points closest to the boundary of interest. Brdiczka et al [12] model using SVM to learn activity labels by first learning the roles of the various entities. Laptev and Linderberg [13] in their work proposed space time interest point operator. Whilst we acknowledge the efforts made through the approaches proposed, researchers have been considering alternative methods for discovering and recognising human activities. Probabilistic topic models inspired by the text and natural language processing community have been applied to discover and recognise human activity routines. Topic model was applied by [14] to discover routines from mobile phone data. Zheng et al [15] proposed the topic model to learn users latent behavioural pattern to discover activities like being at work, going home etc. Huynh et al [3] applied the “bag of words” model of the Latent Dirichlet Allocation (LDA) to discover activities like dinner, commuting, office work etc. This model used the LDA to discover activity labels for proportions of each time window. The major issue with these topic models is that static data proportions have been applied to segments of sensor windows. This means that optimal window sizing has not been considered adequately to discovering certain activities since all activities do not have same duration of occurrence. As an extension of this approach we propose a continuous and dynamic variable segmentation which utilises optimal sensor segments and next most likely activity for continuous activity discovery.

3 Our Proposed Method of Discovering Activities from Sensor Data Observations.

We propose a semi supervised method for discovering activities from continuous sensor data stream as illustrated in Fig 1. The framework we propose through this paper uses partitioned features of sensor data from objects fitted with sensors [4] and then classified using the “bag of word” model of the Probabilistic Latent Semantic Analysis (PLSA).

3.1 Sensors for Activity Recognition in a Home setting.

Objects tagged with state sensors can be used for gathering information about activities undertaken in a home setting. These type of sensors as used by [16] [4] primarily can be in either of the two states of “Open “ and “Close”, “ON” and “OFF” or “0” and “1”. Human activity interactions with objects to which they are attached triggers sensor firings which results to sequence of sensor data made of the described states above depending of the type used. Details of how sequence of sensor data are collected and stored are fully discussed in [17]. A typical sequence of sensor data in this case is a result of sensor firings from activities being performed. Discovering the underlying activities in these sensor events involves partitioning the sequences into segments and then classification of the resulting segments. The challenge therefore is to partition the sequence into the appropriate sizes whilst considering its temporal information for specific activities. We propose a 2 step process to partitioning continuous stream of sensor data – Clustering and then Segmentation.

3.1.1 Clustering Sensor Data Stream.

For an unsupervised clustering, we propose the use of K-means algorithm. Finding the optimal number of clusters is a critical issue in this process. The “Elbow Method” as discussed in [18] can be used to determine the optimal number of clusters which suggests the “number of Topics” or classes. This clustering process assigns to each data point an index which is a measure of its closeness to a cluster of reference. N data points contained in the sensor data stream are clustered such that each sensor data denoted by j is stored to have distances d_j, \dots, N to the centroid of each cluster. We first set $K=N$ and varied the values to $K = 1$ to determine optimal number of clusters. High values of distances from respective centroids invariably meant weak weights. We further used the cluster groupings together with their temporal information as they occur to form segments of sensor sequences.

3.1.2 Segmentation of Sensor Data Stream.

We propose the use of the sensor features from the respective clusters in the section above to build segments of sensor data. We also propose building these segments with considerations

of their temporal information to form time windows for continuous data stream. We have adapted the Variable Sliding Window model (VSW) for the sensor sequence segmentation [19]. To achieve this, we used the sensor features in each cluster as the make up for each segment. We also took consideration of the durations of these segments and the length of their occurrence. With regards to this, the composition of the sensor segments corresponds to the clusters from which they were made but with different lengths in term of their durations. We further calculated the mean duration of each of these segments as given in the equation (1). Let us consider a home fitted with N sensors which are denoted as $x_1 \dots x_n$, occurring in time slices of $t_1 \dots t_n$. The sensor segments are given as $S_1 \dots S_k$ with respect to the total time. If S_i occurs in the time window $t_1 \dots t_n$ for m times then, mean duration of S_i would be:

$$t_{S_i} = \frac{\sum(t_{S_i} - t_{S_i})}{m} \quad (1)$$

For a continuous segmentation process applicable to real time which we propose through this paper, we further considered the next mostly likely segment S_i . The next most likely segment takes into consideration segments co – occurrence and their transitions. A sliding factor ω determines the next segment S_i whilst considering a particular segment S_{i-1} . The sliding factor is a function of time which allows a shift from a segment being considered to the next with or without overlap considerations. We must state however, that the sliding factor is a parameter which we propose to use for defining sensor segment. If a sensor segment S_1 from the sensor data considerations has a very high chance of preceding S_2 and S_2 precedes S_3 , then we slide to next segment from S_1 using the mean duration of S_2 (t_{S_2}) and from S_2 we slide to the next segment using the mean duration of S_3 (t_{S_3}). The estimation of the next most likely segment S_i could be determined by:

$$P(S_i) = P(S_i|S_{i-1}) \text{ where} \\ P(S_1, S_2, \dots, S_k) = \prod_{i=1}^k P(S_i|S_{i-1}) \quad (2)$$

Then the sliding factor ω could be calculated as:

$$\omega = f(t_{S_i} P(S_i)) \quad (3)$$

3.2 Topic Modelling with the Probabilistic Latent Semantic Analysis (PLSA).

The approach we propose follows the “bag of word” framework which was developed from the generative model of the Probabilistic Latent Semantic Analysis (PLSA) introduced by T. Hoffman [2]. The PLSA supposes that there are underlying topics from a document D of words. If D is composed of word sequences $d_1 \dots d_D$, a sequence d_i would be made of words X represented as $x_1 \dots x_n$. The PLSA assumes that a latent topic z from topic $z_1 \dots z_k$ can be classified from $d_1 \dots d_D$ as contained in D i.e for a word x_i contained $x_1 \dots x_n$. In principle, there is a joint probability over $D \times X$ as depicted in Fig. 2 such that a conditional independence assumption that

d and x are independently conditioned on the state of the associated latent topic. This joint probability of d_i, z_i and x_i is given by $P(d_i, z_i, x_i) = P(x_i|z_i) P(z_i|d_i)P(d_i)$ (4)

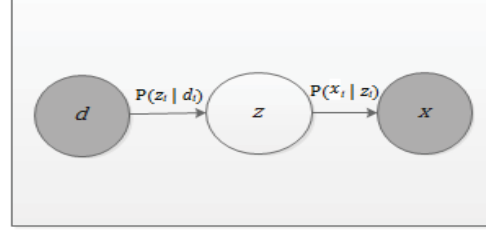


Fig 2. Graphical Model of the Probabilistic Latent Semantic Analysis (PLSA).

The framework makes posterior inference of topic probabilities by the iterative Expectation-Maximisation algorithm. The first step, Expectation (E step) computes the posterior probabilities of the latent variables $P(z_i | d_i)$ from the topic probabilities $P(z_i)$ and the conditional probabilities of the words given the topics $P(x_i | z_i)$. The Maximisation (M step) updates the parameters from the E step by computing the new values for $P(z_i)$ and $P(x_i | z_i)$.

E step: Given $P(x_i | z_i)$ and $P(z_i)$,

$$P(z_i | d_i, x_i) \propto P(x_i | z_i) P(d_i | z_i), \quad (5)$$

M step: Given $P(z_i | d_i, x_i)$ in the E step above,

$$P(x_i | z_i) \propto \sum_{d \in D} n(d_i, x_i) P(z_i | d_i, x_i), \quad (6)$$

$$P(z_i | d_i) \propto \sum_{x \in X} n(d_i, x_i) P(z_i | d_i, x_i), \quad (7)$$

In the context of our activity discovery, the observed word x_i and document d_i of the “bag of word” PLSA corresponds to sensor data observed and sensor segment respectively. Inferred topic z_i then corresponds to activity topic. Intuitively, an activity topic is represented by a probabilistic distribution over observed sensor data events. If we observe $x_1 \dots x_n$ of N sensors in segments $S_1 \dots S_k$ activity topic inference of S_i from this could be computed from $P(z_i|d_i)$ following the EM algorithm given by equation (8).

$$i = \text{argmax}_i P(z_i|d_i) \quad (8)$$

4 Experiments.

To validate our framework we carried 2 some experiments on the Kasteren dataset [4]. The Kasteren dataset was collected using 14 state change sensors over 28 days period. 7 ground truth activities were annotated from the dataset as “Leave House”, “Use Toile”, “Go to Bed”, “Prepare Breakfast”, “Take Shower”, “Prepare Dinner” and “Get Drink”. The first experiment was carried out using 30minutes Fixed Overlap Static Windows FOSW with 3 minutes overlap between segments [20]. For the second experiment, we used K-means explained above to cluster this dataset. We obtained 7 distinct clusters and then further segmented the dataset as explained above. We allowed a 10% overlap of the previously

considered window length. We discovered the activity topics applying the EM algorithm of the PLSA. The activity discovery process was carried out in the MATLAB environment.

4.1 Results

To fully evaluate the performance of our framework we compared our results with the ground truth annotations. A summary of the activities discovered based on our segmentation process are provided in Fig 1. We also calculated the accuracies, precision and F1 from the true positive (TP), false positive (FP), true negative (TN) and false negative of the discovered activities. The accuracy is the measure of how correct our discovered activities are and is calculated by:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (9)$$

The precision is the measure of the relevance of any randomly discovered activity and we calculated this using the equation:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (10)$$

The F1 is the harmonic mean of the precision and is calculated from:

$$F1 = \frac{2TP}{2TP+FP+FN} \quad (11)$$

With regards to the first experiment conducted using FOSW of 30 minutes static segments and 10% overlap. The outcome of the posterior inference for activity topic estimation was misclassification of the posterior probabilities of the activities. This was due to number of observed sensor data in a window segment under consideration. A closer look at the results from the posterior inference, the misclassification was as a result of the lack of consideration of activity duration and the starting times which does not reflect the reality of how the activities took place. For the second experiment, we chose 7 topics as had been estimated by the K means clustering. Our results are reported in Tables 1 and 2 below.

Activities	Instances
<i>Use Toilet</i>	79
<i>Go To Bed</i>	24
<i>Leave House</i>	31
<i>Take Shower</i>	25
<i>Get Drink</i>	5
<i>Prepare Breakfast</i>	10
<i>Prepare Dinner</i>	21
<i>Rejected</i>	6
Total	197

Table 1: Instances of Discovered Activities.

Activities	Accuracy (%)	Precision (%)	F1 (%)
<i>Use Toilet</i>	98.7	100	99.36
<i>Go To Bed</i>	100	100	100
<i>Leave House</i>	100	100	100
<i>Take Shower</i>	100	100	100
<i>Get Drink</i>	66.7	80	80
<i>Prepare Breakfast</i>	46.2	54.5	60
<i>Prepare Dinner</i>	50	52.4	66.7
Average	80.2	83.8	86.6

Table 2: Summary of Discovery Performances.

4.2 Discussion

The results so far presented in the previous sections are indicative of the performance and how robust our proposed framework is for activity discovery. The results from the first experiment resulted in the miscalculation of the posterior inference. We obtained 197 instances of activities discovered (see Table 1) which are close to the groundtruth annotations for the second experiment. For the baseline evaluation and comparisons, we used the groundtruth provided by [4]. The difference are as a result of the varying length of the sensor segments. We also through this process discovered 79 instances of “*Use Toilet*” and 6 instances were “*Rejected*”. The activity discovery results were very high for “*Leave House*”, “*Go To Bed*”, “*Take Shower*” and “*Use Toilet*”. On the other hand, we recorded comparable accuracy and precision results for “*Get Drink*” and “*Prepare Dinner*”. The accuracy for “*Prepare Breakfast*” was 46.2% largely because it was confused with “*Prepare Dinner*” and both activities shared similar object interactions. The “*Rejected*” activity discoveries also accounted for the result we obtained for “*Prepare Breakfast*” and “*Prepare Dinner*”. We believe the overall accuracy we achieved through our proposed framework as 80.2% is significant and appealing.

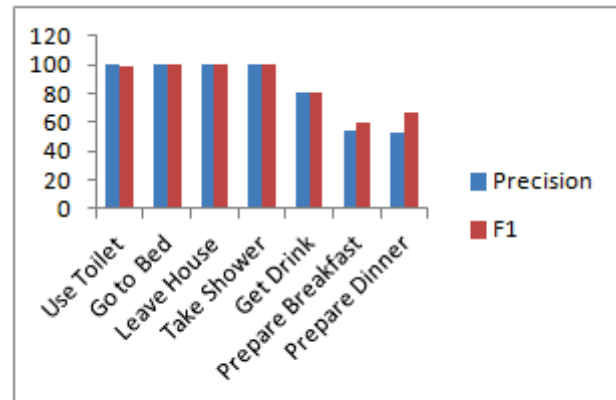


Fig 3. Precision and F1 results of the Activities Discovered.

The Precision and F1 results for the activities are given in Fig 3. The Precision for activity discovery overall was 83.8% indicative of how precise an activity can be discovered through this framework. We discovered through our framework that times covered by as “idle” in the dataset were associated with the other 7 activities. These “idle” times involved object interactions meaning that activities were potentially being carried out. Given these results, we would like to state that the framework we propose is capable of handling realtime activity recognition.

4 Conclusion.

In this paper, we presented activity discovery of sensor data using EM algorithm of the Probabilistic Latent Semantic Analysis (PLSA). This was possible after clustering and partitioning the Kasteren dataset. We turned the this dataset to a “bag of sensor data” documents which corresponded to sensor segments of variable window lengths. The underlying activities were then discovered classified appropriately . Through the framework we propose in this paper, 7 activities given as “Leave House”, “Use Toilet”, “Go to Bed”, “Prepare Breakfast”, “Take Shower”, “Prepare Dinner” and “Get Drink” were discovered. We believe that this approach is appealing given the results achieved. Whilst we have followed this generative approach inspired by the text and natural language processing community, we think it can be exploited to further discover abnormal activities. This could be addressed through the introduction of a layer to discriminate rarely occurring patterns.

Acknowledgements

The authors would like to thank Kasteren et al [4] for their support in the early part of this research and for publicly making their dataset available.

References

- [1] D. Patterson, D. Fox, H. Kautz and M. Philipose, “Fine-grained activity recognition by aggregating abstract object usage,” 2005.
- [2] T. Hofmann, “Probabilistic Latent Semantic Analysis,” Stockholm, 1999.
- [3] H. Tãm, F. Mario and S. Bernt, “Discovery of activity patterns using topic models,” 2008.
- [4] V. K. Tim, N. Athanasios, E. Gwenn and K. Ben, “Accurate Activity Recognition in a Home Setting,” Seoul, Korea, 2008.
- [5] N. Usman, B. John and W. Jinfu, “Recognising Activities of Daily Life Using Hierarchical Plans,” in *European Conference on Smart Sensing and Context*, October, 2007.
- [6] M. T. Emmanuel, S. I. Stephen and L. Kent, “Activity Recognition in the Home Using Simple and Ubiquitous Sensors,” *Pervasive Computing Lecture Notes in Computer Science.*, vol. Volume 3001, 2004.
- [7] A. K. Henry, “A Formal Theory of Plan Recognition and its Implementation,” 1991.
- [8] C. Liming, C. Nugent and W. Hui, “A Knowledge-Driven Approach to Activity Recognition in Smart Homes,” vol. 24, no. 6, June 2012.
- [9] J. Yamato, O. Jun and K. Ishii, “Recognizing human action in time-sequential images using hidden Markov model,” Champaign, IL, 1992.
- [10] N. Oliver, E. Horvitz and A. Garg, “Layered representations for human activity recognition,” 2002.
- [11] M. Brand, N. Oliver and A. Pentland, “Coupled hidden Markov models for complex action recognition,” Media Lab., MIT, Cambridge, MA, USA , 1997.
- [12] B. Oliver, R. Patrick and L. C. James, “Detecting Individual Activities from Video in a Smart Home,” *Knowledge-Based Intelligent Information and Engineering Systems*, vol. 4692, 2007.
- [13] L. Ivan and L. Tony, “Space-time Interest Points,” *IEEE International Conference on Computer Vision*, 2003.
- [14] F. Katayoun and G.-P. Daniel, “Discovering Routines from Large-Scale Human Locations using Probabilistic Topic Models,” vol. 2, no. 1, 2011.
- [15] Z. Jiangchuan and L. M. Ni, “An unsupervised framework for sensing individual and cluster behavior patterns from human mobile data,” 2012.
- [16] N. C. Krishnan and D. J. Cook, “Activity Recognition on Streaming Sensor Data,” vol. 43, no. 3, 2013.

- [17] K. Marc, H. Gerold, F. Alois, C. Alberto, R. Daniel, T. Gerhard, S. Hesam, C. Ricardo, d. R. M. José, B. David, K. Kai and L. Paul, "The OPPORTUNITY Framework and Data Processing Ecosystem for Opportunistic Activity and Context Recognition," 2011.

- [18] S. Sujay, "Large Scale Learning," University of Texas, 2013.

- [19] J. Ortiz, A. Olaya and D. Borrajo, "A Dynamic Sliding Window Approach for Activity Recognition.," Girona, Spain, 2011.

- [20] E. Keogh, S. Chu, D. Hart and M. Pazzani, "An Online Algorithm for Segmenting Time Series," San Jose, CA, USA, 2001.