ACTIVITY RECOGNITION IN A DENSE SENSOR NETWORK

William A. Hoff
Engineering Division
Colorado School of Mines
Golden, Colorado 80401 USA
whoff@mines.edu

James W. Howard
Mathematical and Computer Sciences
Colorado School of Mines
Golden, Colorado 80401 USA
james.w.howard@gmail.com

Abstract

A dense sensor network consisting of passive infrared motion detectors was developed and used to record human activity in hallways and rooms in a large campus building. Algorithms were developed that: (a) automatically determine the topology of the network from the sensor data, so that manual mapping is not required, (b) automatically learn patterns of sensor readings in local spatial and temporal neighborhoods, and (c) use the distribution of local events over larger spatial and temporal scales to automatically discover patterns of underlying global activities. The statistical distribution of the local events is analyzed using Probabilistic Latent Semantic Analysis (pLSA). Preliminary results show that the method can identify "typical" and "anomalous" activities.

Key words: activity recognition, sensor network, motion detectors

1 INTRODUCTION

We address the problem of automatically learning and recognizing activity patterns from a dense sensor network of simple, low cost sensors. We are interested in activities that involve a large number of people acting over a large spatial domain, and possibly over a long period of time. We would like to be able to automatically identify "typical" and "anomalous" activities. This has many important applications such as security in buildings.

A typical approach in security applications is to install numerous cameras and have security guards monitor the video. Even though cameras are relatively cheap, the cost of a security guard to monitor the video is quite expensive. There has been much research on automatically detecting activities from video, but this technology is still in its infancy. There has also been mounting concern over privacy issues in using video in public spaces.

Most current research in activity recognition is limited to activities involving relatively few agents, in a limited area, and over a limited duration. Most current research also requires significant human intervention for modeling and setup. Many researchers use video to provide detailed information on the actions of a single person, such as what they are carrying, their gestures, etc. While we do not rule out the use of video information, we would only use video to derive simple information such as a count of the number of people in the scene.

To avoid privacy concerns and reduce cost, we have developed a dense sensor network composed of simple low cost sensors (passive infrared motion detectors) to monitor public spaces. These sensors can only detect the presence of a person and cannot identify the person, thus alleviating privacy concerns.

Other researchers have also used motion detectors (or similar sensors that are also low resolution and anonymous) to recognize activities. Wilson and Atkeson [1] place sensors in a home and track the movements of an individual, then interpret the activity based on the locations visited. Lymberopoulos et al [2] also recognize activities based on the order of locations that a person visits. However, tracking individual people is not practical in a large building where many people will be present, because of the ambiguity involved. Wu, et al at Intel Research [3], use RFID tags to detect which objects a person touches in the kitchen, then use a dynamic Bayes network to recognize activities. This approach requires a person to wear an RFID reader and extensive tagging of objects or people with RFID tags. A system that was developed for a scenario similar to our own was that reported by Wren, et al [4]. They use a network of motion detectors mounted in the hallways of a building, to recognize activities such as a person going to use the elevator. They require a person to provide both positive and negative training instances of the activity.

In terms of representation, many approaches treat an activity as a temporal sequence of elementary events, which can be represented by a structure such as a hidden Markov model (HMM) [5]. We also use this approach, although we use a simpler fixed model, as described in Section 3.

Our work differs from past work in two important ways: First, we discover activities that involve a large number of people, rather than a single individual. Second, learning is almost completely unsupervised and there is no need for a person to label training instances of activities. We automatically learn patterns of sensor readings in local spatial and temporal neighborhoods. We use the distribution of local events over larger spatial and temporal scales to automatically discover patterns of underlying global activities. This can be used to determine if the observed activity is typical or anomalous compared to previous observations.

2 EXPERIMENTAL DESIGN

A sensor network consisting of 10 wireless motes and 50 passive infrared (PIR) motion sensors was set up in Brown Hall, which is a large building containing offices, research labs, and classrooms. We installed sensors on the ceiling along the hallway on the second floor, in the departmental office, conference room, in an atrium, and at the entrance of classrooms (Figure 1).

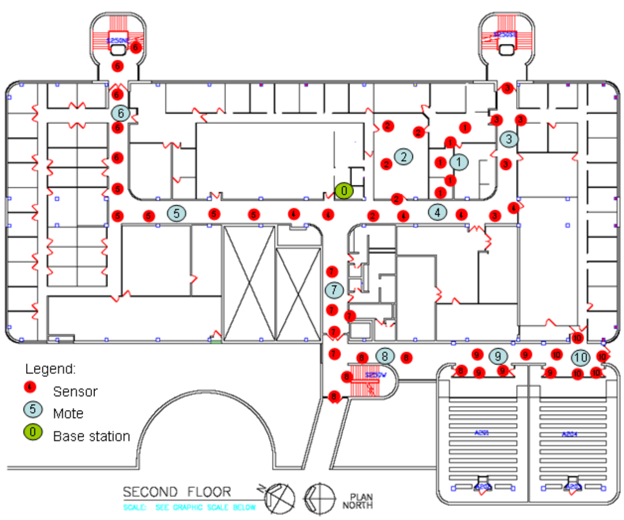

Figure 1 Sensor network in Brown Hall. Small dots are the PIR sensors; the larger dots are the motes. The dot labeled "0" is the network gateway.

The sensors used in this project were the KUBE Electronics PIR Detector Module TR257-1. These sensors are about 1 inch square, and cost approximately \$30-40 each. The sensors detect a moving heat source (such as a person walking by), and are similar to sensors that automatically turn on lights in a room. The sensor is sensitive enough to detect a person is standing under it, and moving slightly. Note that the sensor cannot distinguish between a single person and multiple people. When mounted on an 8 foot ceiling, the sensors have an effective range of about 12 feet.

Sensors were placed rather arbitrarily. We knew the approximate range of the sensors, so we spaced them using that range. Placement was sometimes guided by where we could put them. Sometimes we knew that a lot of traffic would occur in certain places, so we placed more sensors there.

The mote used in this project was the Moteiv Tmote Sky. This mote has an 8MHz microcontroller, 10kB of RAM, 48kB of Flash memory, and runs TinyOS. Each mote costs about \$80 in quantity. It communicates using the 2.4GHz IEEE 802.15.4 (Zigbee) wireless standard. The motes were placed above the ceiling tiles, with wires running to the sensors. The entire sensor network is unobtrusive, with most of the hardware placed out of sight above the ceiling.

Five PIR sensors were connected to each mote, through a custom interface board. We programmed the motes to sample the sensors once every second. If any of the sensors have a detection, then a packet is transmitted to the base station, attached to a PC. An application on the PC server records the detections in a mySQL database. Data was collected continuously from September 2007 to June 2008 (*i.e.*, one academic year), except for a few weeks in which the network was down for maintenance.

One way to visualize the data is in the form of a two dimensional matrix $H(t,s)$, where the columns are sensors and the rows are time intervals. The elements of the matrix are 0 or 1, indicating whether a sensor is active at that time. We actually used an interval of two seconds, since detections are only known to ± 1 second. If any detection occurs within that 2 second interval, it is represented as a hit in the matrix.

Figure 2 shows an example of a sensor hit matrix H. Local movement patterns appear as structures in the matrix. For example, a person walking down the hall causes consecutive hits in adjacent sensors, which appears as a diagonal streak in the data.

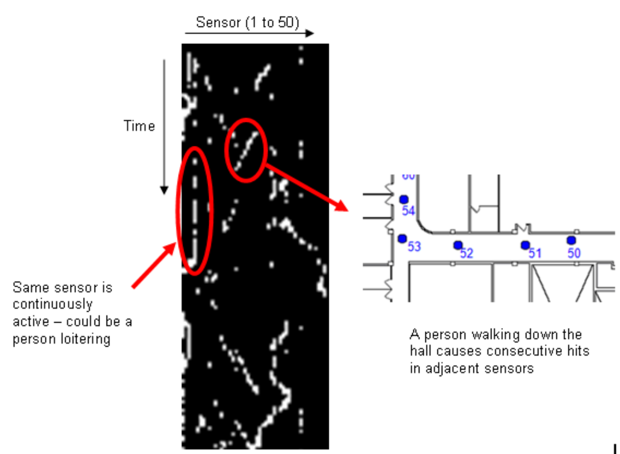

Figure 2 Example of sensor data in the form of a two dimensional matrix.

Of course, the pattern would only take the form of a diagonal streak if the sensors along the hallway are numbered in consecutive order; if the sensors are numbered in a different order, then the pattern might take a different (but predictable) form.

Figure 3 shows the counts for each sensor over a one week period, in the form of "heat map". Brighter colors indicate a larger number of counts. The largest count (in the upper right) is from the sensor over the receptionist's desk in the departmental office.

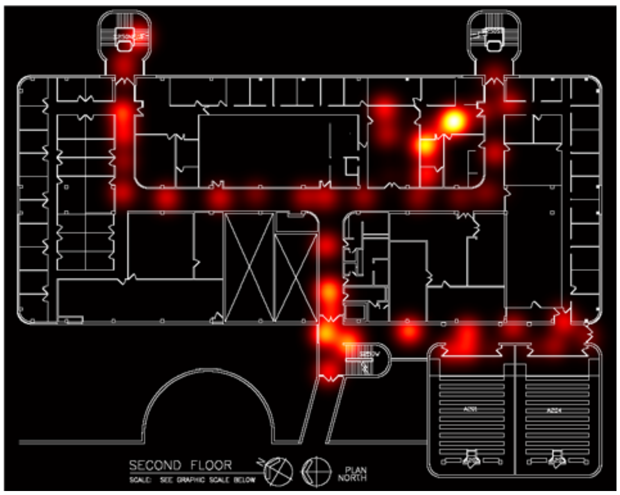

Figure 3 Sensor counts over a one week period.

3 ALGORITHM AND RESULTS

3.1 Network Topology

Our first step is to infer the network topology from the sensor data, similar to the approach of [4]. If two sensors are physically close, and people regularly move from one to the other, there should be a high correlation between the activations as a function of time. We can measure this using the normalized cross correlation coefficient. Let x be activation (0 or 1) of one sensor at time t, and y be the activation of another sensor at time $t+\Delta t$. The normalized cross correlation score is

$$\rho_{x,y} = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$

where $Cov(x, y)$ is the covariance of x and y. The coefficient is equal to 1.0 if the two sensor readings are perfectly correlated, and -1.0 if they are oppositely correlated. It is zero if they are uncorrelated. This can be used to determine the neighbors of each sensor.

Figure 4 shows strong correlations that were experimentally determined using the correlation coefficient. This was determined from readings from a one week period. The thicknesses of the lines between the pairs of sensors correspond to the values of the correlation coefficient between them (only large positive correlations are shown). Correlations were measured for time offsets of Δt=1,2,..,6 seconds. For each pair of sensors, the highest value among all these time offsets was used.

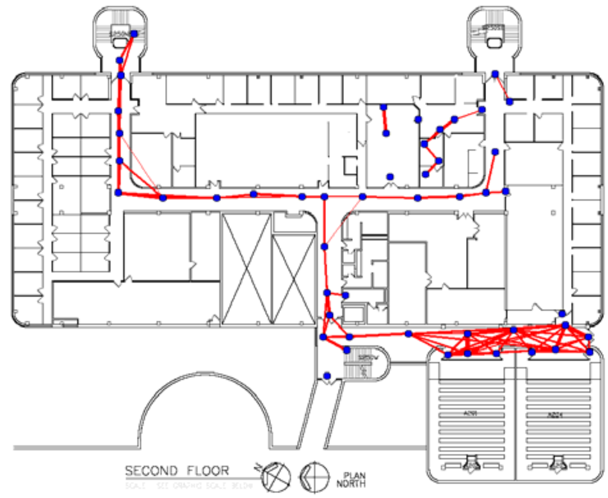

Figure 4 Neighbors of sensors experimentally determined using correlation coefficient.

Sensor pairs with correlation values above a threshold are considered to be neighbors of each other. The threshold was empirically determined to yield neighborhood sizes in the range of 2-5 sensors. In the figure above, the paths of the hallways can easily be seen. This matches the routes where people commonly move through the building. However, sensors that are spatially close, but do not have a physical path between them, are not identified as neighbors.

3.2 Local Activity Patterns

The next step is to detect local, repeated activity patterns. These are localized in space (in terms of the neighborhood of sensors) and time. From the correlation data described in the previous section, each sensor has a neighborhood of size S (consisting of 2 to 5 sensors). Looking over a short time window (T=9), we first find all $H(T,S)$ windows that have a local maximum in the number of counts – these are potentially "interesting" events. As an example, for one such neighborhood, centered on sensor with identification number 50, over one week, there are 1656 such events. Each event X_i is a 9x4 pattern. Figure 5 shows the neighborhood centered on sensor 50, and two candidate events extracted from the data.

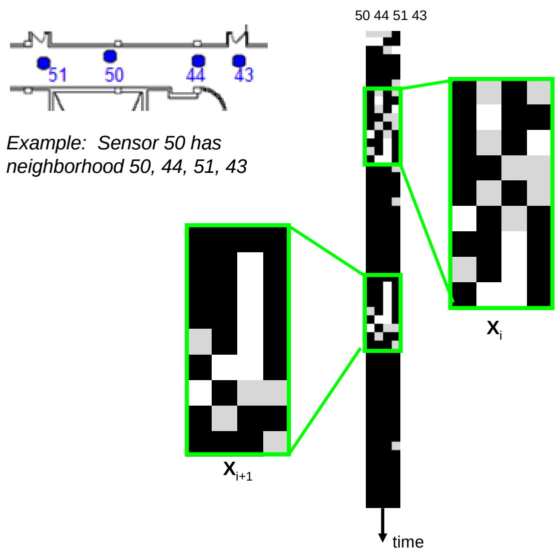

Figure 5 Local patterns with high activity levels are extracted from the data, for each neighborhood.

Next, a k-means algorithm was used to cluster similar events. The normal k-means algorithm had to be modified to deal with the nature of the data. First, the data points are not vectors, but two dimensional matrices of size *TxS*. Second, when computing the distance between a data point and a cluster, the distance measure must allow for potential time shifts, caused by a slightly different starting point. The distance measure between events X_i and X_j is given in the equation below.

$$d_{i,j} = \min_{dt}\left(\sum_{t=1}^{T} \sum_{s=1}^{S} \left(X_i(t+dt,s) - X_j(t,s) \right)^2 \right)$$

We also re-centered the clusters in time after each iteration. With these modifications, the k-means algorithm does not converge monotonically, in the sense that the sum of the point-to-cluster distances always decreases with each iteration. However, the observed behavior of the algorithm was that the sum of distances decreased rapidly and then fluctuated slightly about a low value. We terminated the algorithm after a fixed number of iterations (20).

We used a value of k=7 for the desired number of clusters (chosen empirically). We discarded clusters that had too few points assigned to them. The result was that we found between 2 and 7 clusters for each neighborhood.

The cluster results for neighborhood 50 are shown in Figure 6. Looking at the meaning of these clusters, we can interpret C1 as a person walking to the left in the hallway. C2 and C3 represent a person walking to the right (at slightly different speeds). Preliminary tests show that the patterns with the highest counts are repeatable; i.e., very similar from one week to the next.

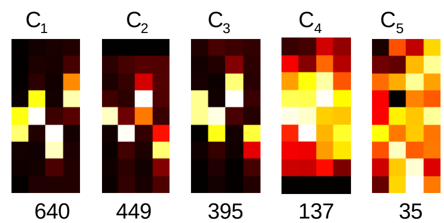

Figure 6 Cluster results for the neighborhood centered at sensor 50, with number of occurrences in each cluster.

These clusters represent localized, repeated activity patterns. We then identify the occurrences of these patterns in the entire data set, using cross-correlation. The entire data set is thus converted to a sequence of symbols, each of which represents the occurrence of a local pattern. We can then analyze the distribution of symbols over time periods, and the dynamic sequences of symbols. This is described in the next section.

3.3 Analysis of Event Distribution

Our approach for analyzing the distribution of local events over time was inspired by the method of probabilistic latent semantic analysis [6]. The pLSA method was originally developed for document analysis. It can discover a small number of latent or "hidden" topics in a large collection of documents. A document is described using a word frequency vector, containing the frequency of each word w_i in the document d_j. It uses an iterative expectation-maximization (EM) algorithm to discover a set of underlying topic vectors, or latent classes z_k. Given any new document, one can estimate the composition of the document in terms of its latent classes.

In our application, "words" are the local activity patterns found by clustering. There were 291 total unique symbols, each representing a particular pattern at a particular sensor neighborhood. The equivalent of "documents" is sensor data recorded over a specific time interval (such as hours or days). We performed a pLSA analysis of sensor data from 17 sensors along the hallway. Data from one month was analyzed.

Data from a one-month period (March 1-30) was divided into days. Each "document" is a day, so there are 30 documents. We specified that there were 8 latent classes. Looking at the results of the analysis, the composition of "topics" for each document appears to be dependent on the day. In particular, latent class 2 seems to correlate with "weekday", and latent class 4 seems to correlate with "weekend" or "holiday". Figure 8 shows the probability (vertical) of a latent class for each day of the month (horizontal axis). The plot shows that class 2 is high during weekdays (M-F) and class 4 is low on

Saturdays and Sundays (Mar 1-2, 8-9, 15-16, 22-23). Interestingly, March 10-14 was spring break, and this period (circled) showed characteristics of both weekdays (class 2) and also weekends (class 4).

To get some insight into the composition of the latent classes, we can look at the probability that a symbol (or activity pattern) occurs within each latent class. There are a large number (291) of these symbols, so we only look the ones with the highest probability.

For class 2 ("weekday"), two of the most common patterns are centered on sensor 72. These patterns are shown in Figure 7. Both appear to show a person loitering; one for a short time and the other for a longer time. Since these sensors are over the janitor's closet and outside the restroom, this possibly could indicate the janitor mopping.

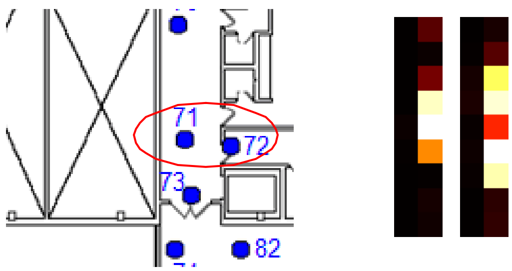

Figure 7 The highest probability patterns comprising latent class 2.

We can also analyze how the days ("documents") differ from each other, in terms of the latent classes. Figure 9 is a plot of each day, according to its proportion of latent class 2 (horizontal axis) and latent class 4 (vertical axis). Note that weekdays cluster in the upper right. Sunday March 30 is an anomaly – it has an unusually large value for latent class 4, and does not group with the other Saturdays and Sundays. We looked at the raw data and it appears that a large meeting or gathering occurred in the atrium that afternoon

4 CONCLUSIONS AND FUTURE WORK

We have developed methods to represent and recognize group activities from dense sensor network data. Advantages of the method are that very little *a priori* information is used – learning is almost completely unsupervised. First, sensor data is used to infer topology of the network. Next, a simple k-means clustering algorithm discovers repetitive patterns, which are localized in space and time. More robust clustering algorithms, that attempt to choose better initial cluster locations, or escape local maxima, could be investigated.

The data is represented by discrete symbols, corresponding to the discovered local patterns. The statistical distribution of the symbols is analyzed using Probabilistic Latent Semantic Analysis (pLSA). Preliminary results show that the method can identify "typical" and "anomalous" activities. Possible applications of this method include building security, and prediction of human activity for the purpose of improving building energy efficiency.

Currently, our simple PIR sensors limit the types of activities that can be recognized. For example, the sensors are incapable of identifying or tracking individual people, or accurately knowing the number of people in a given area. We are looking at adding other types of sensors, including sound level, light intensity, and sonar, which may provide additional information.

5 ACKNOWLEDGEMENTS

We acknowledge the support of the National Science Foundation (grant CNS-0931748) and the Lockheed Martin Corp., and contributions by Michael Colagrosso, Nalinrat Guba, Bao Nguyen, and Philip Loden.

6 REFERENCES

[1] D. Wilson, C. Atkeson, "Simultaneous tracking and activity recognition (star) using many anonymous, binary sensors," 3rd International Conference on Pervasive Computing, pp. 62–79, 2005.

[2] D. Lymberopoulos, et al, "Detecting Patterns for Assisted Living Using Sensor Networks," Proc. of SensorComm, Valencia, Spain, October 2007.

[3] J. Wu, et al, "A Scalable Approach to Activity Recognition based on Object Use," ICCV 2007, Rio de Janeiro, Brazil, October 2007.

[4] C. Wren, et al, "Similarity-based analysis for large networks of ultra-low resolution sensors," Pattern Recognition vol. 39, pp. 1918-1931, 2006.

[5] N. Oliver, et al, "Layered Representations for Learning and Inferring Office Activity from Multiple Sensory Channels," Computer Vision and Image Understanding, vol. 96, pp. 163-180, 2004.

[6] Hofmann, T., "Probabilistic Latent Semantic Indexing", Proc. of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval, 1999.

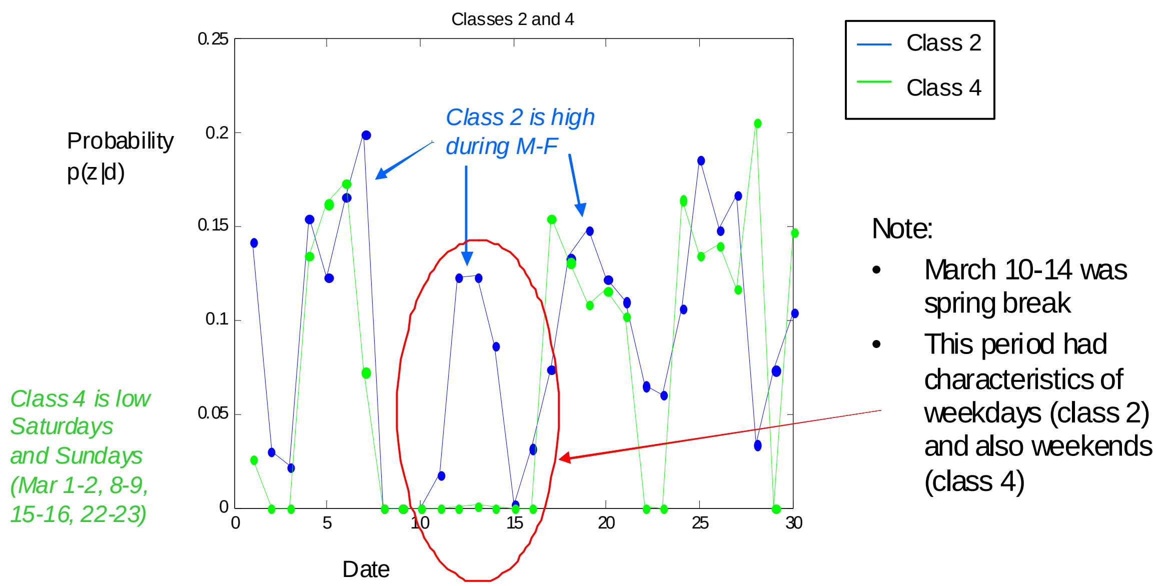

Figure 8 Probability of latent classes 2 and 4, for the analysis where a "document" is one day.

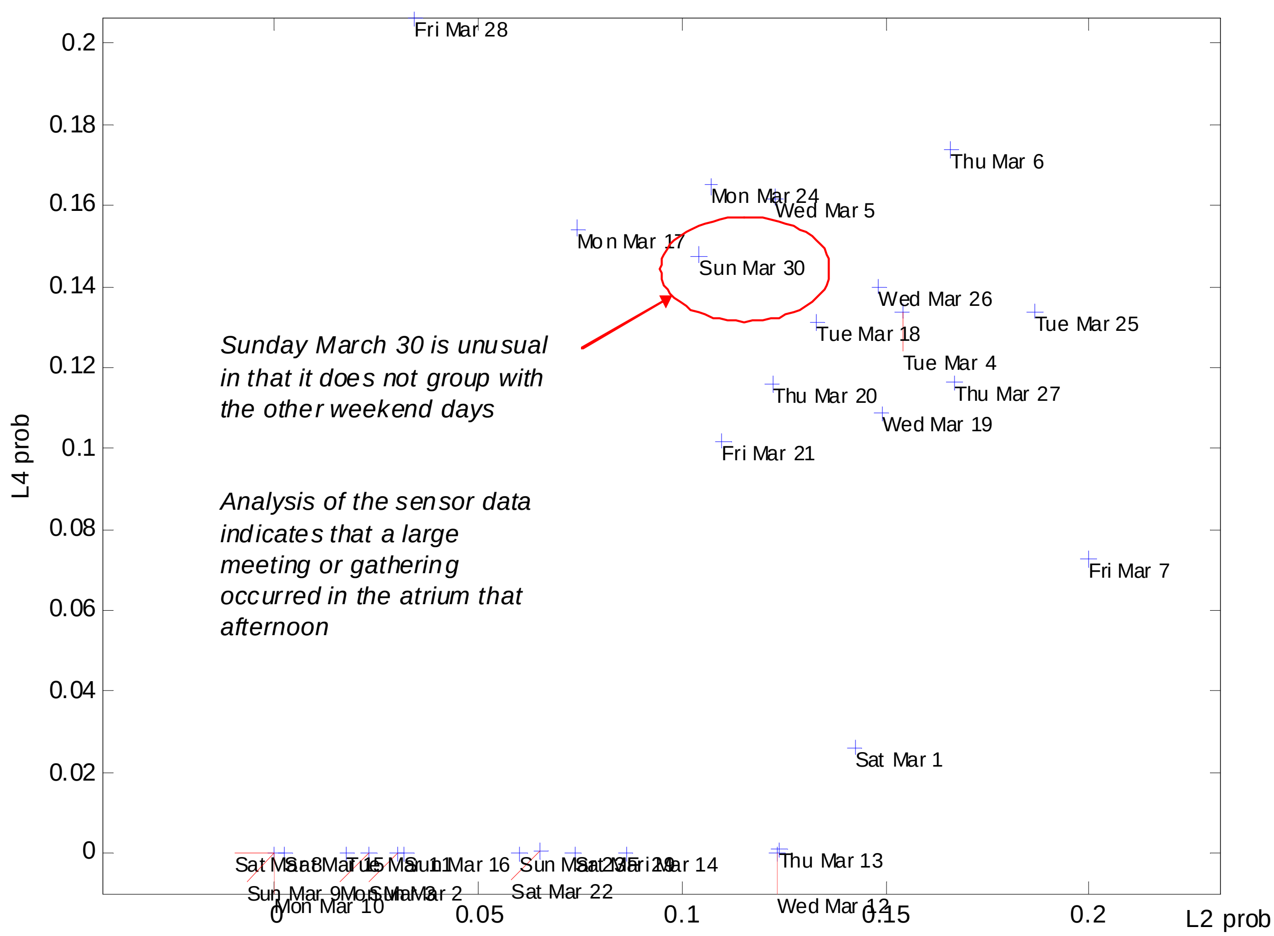

Figure 9 Plot of each day in terms of probabilities for latent classes 2 and 4