

Quantitative and Qualitative Comparison of Periodic Sensor Data

Akshay Jain, *Student Member, IEEE*, James M Keller, *Life Fellow, IEEE* and James C Bezdek, *Life Fellow, IEEE*

Abstract— The recent rise in the availability of sensors to monitor our day-to-day lifestyles calls for techniques to glean information out of the generated data. For many of these sensors, the data recorded is periodic in nature. Moreover, it has been shown that a change in this periodic data is correlated to a shift in physical or mental health. We present a technique to compare this repetitive data, both qualitatively and quantitatively. Our method first summarizes the sensor measurements collected from smart homes of elderly into textual statements. Dissimilarity between the text summaries is then computed to quantify the comparison.

I. INTRODUCTION

In the last few years, continuous monitoring of our lifestyle using sensors has gained momentum. This includes, but is not limited to, commercially available devices such as Fitbit, which records our walking habits, smart watches which along with the motion, can measure heartbeat, and a myriad of smart phone apps which make use of accelerometers and gyroscopes present in mobile devices. Another non-intrusive method to monitor people's activities are networks of sensors installed in 'smart homes'. These devices include motion sensors to monitor activities inside homes or sensors installed in bed to monitor sleeping patterns [1]. All of these sensors, monitor the daily/weekly/monthly routines of people using them. For example, a person might have similar walking patterns throughout a week or similar patterns for all the weekdays but different on the weekends. It has been shown that changes in certain activity patterns may signal changes in person's physical or mental health condition [2]. Recognizing these changes ahead of time may help in preventing forthcoming health problems.

A common way to understand and extract information from this data is to visualize it using graphs or some kind of display. However, this requires a viewer to carefully look at these displays and make decisions based on them. A more independent way would be a method that monitors the data, and triggers an alarm that signals a change autonomously. An alarm should only be raised depending on the degree of change in the patterns. For example, when monitoring weekly activity, a change in one day's activity shouldn't trigger an alarm, but if the activity over the whole week starts changing, then it might be a cause of concern.

Several studies to test the hypothesis that health problems lead to changes in sensor data pattern was conducted in [2]. In this work the authors showed that mental illness such as depression and dementia influences the patterns recorded by sensors monitoring in-home activity and the amount of time

residents spend outside the apartments. In this work they analyzed this data visually and pointed at the difference in the sensor data before, during and after a mental illness. We take inspiration from this work and design a method that attempts to find these differences between the patterns of activity autonomously. We use Linguistic Protoform Summaries (LPS) as a way to extract features of data and present an approach to work with these features. We propose a method to describe several days' worth of data linguistically, along with a technique to compare these linguistic descriptions quantitatively.

II. BACKGROUND

Our data comes from an Aging in Place facility called TigerPlace in Columbia, Missouri. The apartments at TigerPlace are equipped motion sensors, bed sensors and Microsoft Kinect to monitor the activities of elderly. The focus of this work is on the network of motion sensors which are carefully placed at different positions in the kitchen, bedroom, living room etc. to monitor activity all around the apartment. The data from these sensors is sent to a remote database over a wireless internet connection. This data can then be visualized over a web interface, which among other things, shows the motion around the apartment and out of apartment visits in the form of a density map. Figure 1 presents one such density map representing activity around an apartment for a period of 30 days. In the figure, each column represents a day, while each row is one hour of a day starting from 12:00 AM through 11 PM. The colored squares represent the motion density, which is the aggregate of the number of sensor hits around the apartment in that hour divided by the fraction of time the resident spends inside the apartment. The black patches indicate the times the person was not present inside the apartment, computed by a separate algorithm from the motion sensor data.

In [2] the authors show that the density map of a month when the elder resident was diagnosed with depression and/or dementia was very different from the maps of a month before the diagnosis and after taking prescribed medications. In that work, five case studies were presented, two of which are shown in Figures 3 & 4.

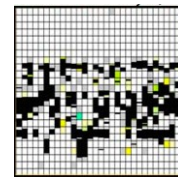


Figure 1: Density map example

Akshay Jain is supported by AHRQ Grant 1R01HS023328 and Center for Eldercare and Rehabilitation Technology at the University of Missouri. Akshay Jain (e-mail: aj4g2@mail.missouri.edu), James M Keller (e-mail:

kellerj@missouri.edu) and James C Bezdek (jcbezdek@gmail.com) are with Department of Electrical and Computer Engineering at the University of Missouri, Columbia, MO 65211 USA.

Using the density maps, they observe that in most of the cases, during mental illness, people tend to stay inside the apartment, but with less activity and go outside less frequently. An approach to compare density maps was presented in [3] where textural features were used to find dissimilarity between density map images. In this work we take two of those case studies and show that our method using LPS can distinguish the 3 density maps linguistically (before, after and during mental illness) as well as compare them quantitatively.

Linguistic Protoform Summaries are short template-based sentences providing a snapshot of data in textual form. In the sense of Yager [4], they are comprised of a Quantifier (*Most, Few* etc.) conveying information about the quantity of the attribute being summarized, a Summarizer (*long, short* etc.) measuring the feature of concern and a truth value (0 to 1) which signifies the validity of the summary with respect to the data. For example, a summary focusing on out of apartment visits during a week might look like, *Most of the out of apartment visits during this week were long*. In the following, we describe a method to find such LPS.

LPS have been utilized in the past to express sensor data from apartments of elderly [5, 6]. In [5] we used LPS to describe sensor data leading to health alerts in a Natural Language Generation setting. We used LPS as an engine to extract useful information from data and then defined some language rules to make the LPS more intuitive. A method to compute distance between two LPS was presented in [7] with a proof that this distance is a metric. This method was then employed in [8] to generate alerts in eldercare data using LPS. This was further enhanced in [9] to produce Linguistic Medoid Prototypes (LMP). LPS were used to describe each night of sleep activity, which were then clustered to find the most prominent LPS representing all of the nights.

Following the work in [7] and [9], we to compute LPS representing the density maps and also compare them quantitatively. We show that LPS has a potential to describe and compare data in both qualitative and quantitative fashion.

I. METHOD

In this section we describe our technique to compare the density maps mentioned in [1]. Two such cases are shown in Figure 3 and 4. The density map in the center of each figure shows the activity of a resident during a month when he/she was diagnosed with depression and dementia. The image to the left shows the activity before being diagnosed and the one on the right shows the activity after taking medications for treating the depression and/or dementia. Visually, we see that during the month when the resident was suffering from the mental illness, the activity changed drastically. Even though there is some change in the motion activity inside the apartment, the most significant difference is in the decrease in out of apartment activity (OOA), which is what we focus on in this work. However, performing such a comparison using some computerized algorithm is not a trivial task. We describe a method using Linguistic Protoform Summaries (LPS) in the following.

A. Linguistic Protoform Summaries (LPS)

In order to compare the density maps using LPS, we first need to describe each map linguistically. The foremost step is to define a protoform that completely describes the OOA activity of a day.

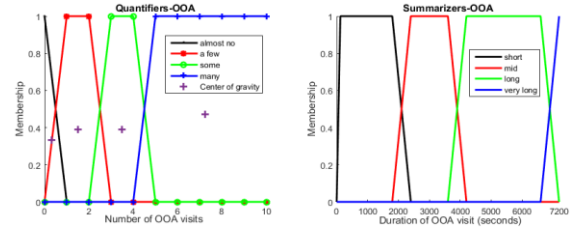


Figure 2: Quantifier and Summarizer Membership functions. The CoG for Quantifiers are also shown.

We use the protoform:

The resident made Q S out of apartment visits.

Basically, S is used to measure the length of the OOA activities while Q quantifies how many of such activities took place. Formally, Q and S are the quantifier and summarizer, respectively, modelled by fuzzy sets as shown in Figure 2. Next, we find a set of expressions which define the OOA activity of any given day. This is done by finding a quantifier for each summarizer, which is the one with the highest truth value, a departure from the approach in [8, 9]. That is, for each summarizer we can have four expressions, one with each quantifier. We only take the one which has highest truth value. The truth value is computed using the method presented in [10], as shown in Equation (1).

$$T(A \text{ y's are } P) = \max_{\alpha \in [0,1]} (\alpha \wedge A(P_\alpha)) \quad (1)$$

where \wedge is the minimum operator, $P(x)$ is the membership function of the summarizer P , $P_\alpha = \{y_i \in Y \mid P(y_i) \geq \alpha\}$ is the proportion of objects whose membership in $P(x)$ is greater than or equal to α (varies from 0 to 1 in small intervals), $|\cdot|$ denotes the cardinality of a set and $A(x)$ is a normal, convex and monotonically non-decreasing membership function of the quantifier A . For a quantifier whose membership function is not monotonically non-decreasing, it is split into two monotonically non-decreasing functions, $A_1(x)$ and $A_2(x)$ (which is used to compute $A_2(x)$) and the truth value is computed as shown in (1). Please refer [10] for more details. Using this method, day one of the density map in the left of the Figure 3 is described as shown in Table 1.

Once we have linguistic description of each day, we use single linkage hierarchical clustering to cluster the LPS/days to find LPS representing a density map. But first, we need to define a method to compare two LPS sets quantitatively.

Table 1: LPS to describe out of apartment activity during a day

Quantifier (Q)	Summarizer(S)	Truth Value
<i>few</i>	<i>short</i>	1
<i>few</i>	<i>mid</i>	1
<i>almost no</i>	<i>long</i>	0.79
<i>few very</i>	<i>long</i>	1

B. Quantitative Comparison two LPS sets

Consider two LPS sets L and M with N_S summaries each, (N_S being the number of summarizers). The dissimilarity between the two sets is given by Equation (2)

$$dissim(L, M) = \left(\frac{\sum_{i=1}^{N_S} (1 - sim(l_i, m_i)) W_i}{\sum_{i=1}^{N_S} W_i} \right) \quad (2)$$

Here,

$$sim(l_i, m_i) = sim(Q_1 \text{ y's are } P, Q_2 \text{ y's are } P)$$

$$= \min(sim(Q_1, Q_2), sim(T_1, T_2))$$

is the similarity between individual summaries l_i and m_i that have a common summarizer. The similarity between the truth values is computed as shown below.

$$sim(T_1, T_2) = 1 - (T_1 - T_2)$$

The similarity between Quantifier membership functions, Q_1 and Q_2 is computed using a method in [11] and is given by Equation (3).

$$sim(Q_1, Q_2) = 1 - \frac{\sum_{i=1}^4 |a_i - b_i|}{4} \quad (3)$$

where a_i and b_i are defining points of trapezoidal membership functions, Q_1 and Q_2 (membership functions in Figure 2).

In (2), $W_i = \max(W_{l_i}, W_{m_i})$, where W_{l_i} and W_{m_i} are the weights associated with the summaries l_i and m_i , respectively. These weights are the x coordinate of the center of gravity of the quantifier membership functions, as defined in [11]. They are used to encode the reasoning that while comparing two summaries, the dissimilarity between them should be assigned more weight if they collectively represent more objects as compared to the pair which represent a lesser number of objects. For example, if the quantifier associated with both l_i and m_i is *almost no*, this should have lesser weight than if the quantifiers are *many* and *some*. The centers of gravity for the quantifier membership functions are shown in Figure 2.

Using this method to compare sets of LPS representing each day in a density map, as shown for example in Figure 1, we can now cluster all the LPS sets using a clustering algorithm such as Single Linkage Clustering. Each density map is represented by the set of clusters thus found. The prototypes of these clusters can then be said to represent the complete month of out-of-apartment activity. We call these prototypes Linguistic Medoid Prototypes (LMP). For each cluster, the LMP is the LPS set which is at minimum distance to all the objects in that cluster. In other words, the prototype of a cluster represent the day that is most similar to all of the LPS sets in that cluster. Each prototype is accompanied by the number of elements/objects in its cluster, called the Support of the prototype.

This method is employed to compute LMP for all the density maps. To compare one density map to another, we now need a method to compare two sets of LMP, that is, LMP representing the two density maps.

Let A and B denote the sets of LMP (from the clustering) representing two density maps with the individual cluster prototypes denoted by L^i and M^j respectively for each set.

With N_A and N_B being the number of prototypes (clusters) in each set, the dissimilarity between A and B is given by

$$dissim(A, B) = \frac{\sum_{i=1}^{N_A} W_A^i (\min_{j=1, \dots, N_B} dissim(L^i, M^j))}{\sum_{i=1}^{N_A} W_A^i + \sum_{j=1}^{N_B} W_B^j} + \frac{\sum_{j=1}^{N_B} W_B^j (\min_{i=1, \dots, N_A} dissim(L^i, M^j))}{\sum_{i=1}^{N_A} W_A^i + \sum_{j=1}^{N_B} W_B^j} \quad (4)$$

Here, the dissimilarity between L^i and M^j is computed using (2). Basically, for each LPS set in the first prototype set, we find the closest LPS set in second set. Then the same procedure is carried out for the second set where we find the closest LPS set in first group. The final dissimilarity between A and B is the weighted average of all dissimilarities. The weights W_A^i and W_B^j are the weights assigned to the dissimilarities when comparing LMP of set A with set B and the other way round, respectively. The weight is the total support involved in a comparison. For example, while comparing summary L^1 of A to the prototype set B, let's say the closest prototype in B is found to be M^2 . Then, the weight assigned to this comparison is the sum of the support of L^1 and that of M^2 .

We applied this method to compare the OOA activity in the two sets of density maps of [2] as shown in Figures 3 and 4. As mentioned before, for each case, the density maps in the center are for a time period when the resident was diagnosed with depression, the one on the left is for the period before the diagnosis and the one on the right is for a period after taking medications. Also, it should be noted that each density map contains resident's activity for approximately one month (that is, it's not exactly a month).

II. RESULTS

In this Section we show the results obtained by applying the method mentioned in Section 3 to two of the cases of [2]. For each density map, while performing clustering, the number of clusters were set to 3. After the clustering, the clusters with support less than 3 were rejected, since they do not impart significant information to the results. Also, note that the days with no OOA activity are represented by the LPS set with *almost no* as the quantifier accompanied by each summarizer and with a truth value of 1.

A. Case Studies

Figure 3 and 4 shows density maps for two residents. For both of these cases, we can see visually that the density maps for the months before and after mental illness look very different from the month when the person was suffering from illness. This can also be observed from the linguistic prototypes beneath the density maps. The prototypes for both before and after mental illness time periods describes a lot of out of apartment visits with varying duration. However, while suffering from mental illness he/she tends to stay inside the

apartment which is denoted in the linguistic prototype with *almost no out of apartment visits* as the only LMP. The dissimilarity between the linguistic prototypes as computed by (4) is shown in Table 2. It depicts the visual and linguistic comparison between the density maps quantitatively. For both cases, we observe that the dissimilarity between before and after mental illness is less than the dissimilarity between other pair of density maps.

Table 2: Dissimilarity between the prototypes

	Dissimilarity between - prototypes		
	Before- During	During- After	Before- After
Case 1	0.179	0.215	0.097
Case 2	0.129	0.156	0.089

III. CONCLUSION

With the help of two case studies we showed that the presented method can describe periodic sensor data linguistically as well as compare it quantitatively. The method is still in early stages but the results look promising. In this work, we only focused on the out of apartment data. To see if the method scales well, other information such as motion density should also be used. An important aspect of this method is the choice of protoforms, which should be carefully chosen for different types of data. Finally, to better assess the method, evaluation with more case studies is required.

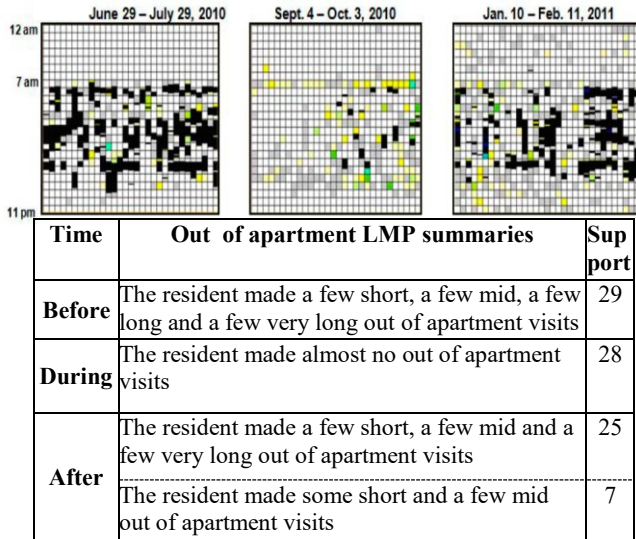


Figure 3: Density maps for resident case # 1 with the linguistic prototypes for the month before, during and after mental illness.

REFERENCES

- [1] M. J. Rantz, M. Skubic, S. J. Miller, C. Galambos, G. Alexander, J. Keller, and M. Popescu, "Sensor technology to support aging in place," *Journal of the American Medical Directors Association*, vol. 14, pp. 386-391, 2013.
- [2] C. Galambos, M. Skubic, S. Wang, and M. Rantz, "Management of dementia and depression utilizing in-home passive sensor data," *Gerontechnology: international journal on the fundamental aspects of technology to serve the ageing society*, vol. 11, p. 457, 2013.
- [3] S. Wang, M. Skubic, and Y. Zhu, "Activity density map visualization and dissimilarity comparison for eldercare monitoring," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 16, pp. 607-614, 2012.
- [4] R. R. Yager, "A new approach to the summarization of data," *Information Sciences*, vol. 28, pp. 69-86, 1982.
- [5] A. Jain and J. M. Keller, "Textual summarization of events leading to health alerts," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015, pp. 7634-7637.
- [6] A. Wilbik, J. M. Keller, and G. L. Alexander, "Linguistic summarization of sensor data for eldercare," in *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, 2011, pp. 2595-2599.
- [7] A. Wilbik and J. M. Keller, "A distance metric for a space of linguistic summaries," *Fuzzy Sets and Systems*, vol. 208, pp. 79-94, 2012.
- [8] A. Wilbik and J. M. Keller, "Anomaly detection from linguistic summaries," in *Fuzzy Systems (FUZZ), 2013 IEEE International Conference on*, 2013, pp. 1-7.
- [9] A. Wilbik, J. M. Keller, and J. C. Bezdek, "Linguistic Prototypes for Data From Eldercare Residents," *Fuzzy Systems, IEEE Transactions on*, vol. 22, pp. 110-123, 2014.
- [10] A. Jain and J. M. Keller, "On the computation of semantically ordered truth values of linguistic protoform summaries," in *Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on*, 2015, pp. 1-8.
- [11] S.-J. Chen and S.-M. Chen, "Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers," *Fuzzy Systems, IEEE Transactions on*, vol. 11, pp. 45-56, 2003.

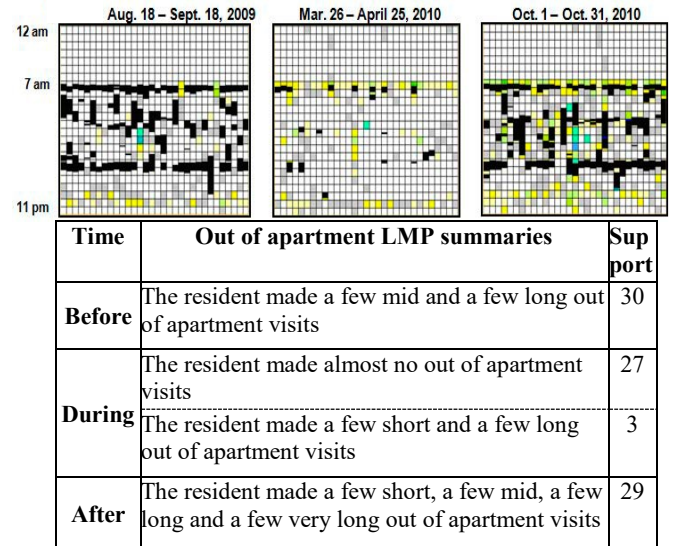


Figure 4: Density maps for resident case # 2 with the linguistic prototypes for the month before, during and after mental illness.