Monitoring Household Electricity Consumption Behavior for Mining Changes

Christian Nordahl*, **Veselka Boeva**, **Håkan Grahn** and **Marie Persson Netz**
Blekinge Institute of Technology
{christian.nordahl, veselka.boeva, hakan.grahn, marie.netz}@bth.se

Abstract

In this paper, we present an ongoing work on using a household electricity consumption behavior model for recognizing changes in sleep patterns. The work is inspired by recent studies in neuroscience revealing an association between dementia and sleep disorders and more particularly, supporting the hypothesis that insomnia may be a predictor for dementia in older adults. Our approach initially creates a clustering model of normal electricity consumption behavior of the household by using historical data. Then we build a new clustering model on a new set of electricity consumption data collected over a predefined time period and compare the existing model with the built new electricity consumption behavior model. If a discrepancy between the two clustering models is discovered a further analysis of the current electricity consumption behavior is conducted in order to investigate whether this discrepancy is associated with alterations in the resident's sleep patterns. The approach is studied and initially evaluated on electricity consumption data collected from a single randomly selected anonymous household. The obtained results show that our approach is robust to mining changes in the resident daily routines by monitoring and analyzing their electricity consumption behavior model.

1 Motivation and State of the Art

In this study, we develop a model that can be applied for monitoring household electricity consumption behavior and mining changes that can eventually be used to recognize shifts in routines and sleeping patterns of the resident, such as sleep disturbances.

Recent studies in neuroscience suggest a link between dementia and sleep disorders [Brzecka *et al.*, 2018; Cirpiani *et al.*, 2015; de Almondes *et al.*, 2016]. Insomnia and other sleep disturbances are common in patients with neurodegenerative disorders, such Alzheimer's disease and other dementing disorders [Cirpiani *et al.*, 2015]. Changes in sleep of patients with Alzheimer's disease are often observed on very early stage, e.g., a usual 20-minute daytime nap transforms into several hours per day. Sleep disorders can be an important diagnostic indication that foreruns development of Alzheimer's disease pathological disorders. A systematic review published in [de Almondes *et al.*, 2016] provides data supporting that insomnia may be a predictor for dementia in older adults.

Recently, sleep monitoring based on off-the-shelf mobile and wearable devices has emerged as a way to obtain information about one's sleeping patterns [Shelgikar *et al.*, 2016], [Ping-Ru *et al.*, 2015]. By taking advantage of diverse sensors, behaviors and routines associated with sleeping can be captured and modelled. For example, in [Buet, 2017] the bio-signals of a pool of Parkinson's disease patients collected from ambient sensors have been analyzed to detect sleep disorders. What makes sensor monitoring particularly attractive is the non-invasive nature of the sensing compared for example, to traditional Polysomnography. Similarly, with the adoption of smart meters in the electrical power grids, we can now collect high resolution electricity consumption data remotely on a household level. This type of data can be used to get insight into the residents' habits and activities, with low impact and intrusion of the residents' privacy compared to sensor data. As it was discussed above, dementia and other neurodegenerative diseases can cause changes in the behavior of the individual by provoking insomnia, apathy, restlessness etc. We believe that such changes in an individual's daily behavior, can be caught by their electricity consumption activities.

In our previous work, we have applied clustering techniques for analyzing and understanding households' electricity consumption data [Nordahl *et al.*, 2018]. The knowledge extracted from this analysis was then used to create a model of normal electricity consumption behavior for each particular household, which was later described in [Nordahl *et al.*, 2019]. In this work, we propose and investigate an approach that can be used to monitor such a model over time, detect changes and further analyze them on their meaningfulness. The proposed approach builds a new model on a new set of electricity consumption data collected over a predefined time period and then compares the existing historical model with the new household electricity consumption behavior model. If a discrepancy is discovered we perform a further analysis of the current electricity consumption behavior of the elderly habitant in order to investigate whether this discrepancy is associated with alterations in their daily routines. The proposed approach can also be applied to monitor and look for alterations in the sleep-wake cycle of elderly individuals who have already been diagnosed with dementia in order to avoid risk of falling, and identify the need for nursing home placement.

*Contact Author

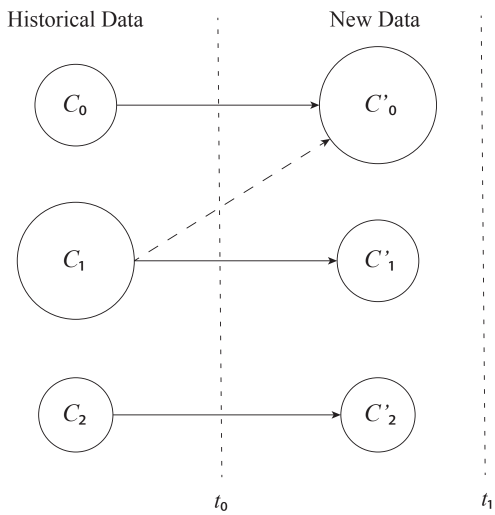

Figure 1: Schematic illustration of the proposed approach.

2 The Proposed Approach and Methods

In order to model the ordinary (normal) electricity consumption behavior of a household, we create k number of clusters with the use of k-medoids. The medoids of these clusters are then defined to model the normal electricity consumption behavior of the resident. We do not have information about the number of people living in the household, therefore we assume the resident lives alone. These signatures (the cluster medoids) are then used as a baseline to compare against for a new portion of data that arrives and also used as initial seeds for the clustering algorithm that is applied on the new data. The reason for using the existing cluster medoids as the initial seeds for the new data is the assumption that the new data will support the same electricity consumption behavior model as the old one. In addition, this enables the analysis and tracking of changes in the existing clusters and further detection of alternations in the electricity consumption behavior modes of the household. For example, we might trace cluster evolution through the detection of transitions, such as cluster shrinking, merging, splitting etc. This idea is schematically illustrated in Figure 1 by a bipartite graph, where the existing clusters (generated on historical data) and the newly generated clusters (used a new portion of data) are sets of left and right nodes and the arrows represents transition correlations between the clusters of two solutions.

The new portion of data is also clustered with k-medoids, but as mentioned above the clustering is initialized by the medoids of the existing clustering solution. After the initial partitioning of the data into the clusters, the initial cluster medoids are removed from the data and new medoids are selected, then k-medoids iteratively continues to refine the clustering until it completes. We then compare the two clustering solutions (the existing and the new one) by calculating the Dynamic Time Warping (DTW) distance between their corresponding sets of medoids. If there is a discrepancy between the two clustering solutions, e.g., the calculated DTW is above the pre-defined threshold, a change in the electricity consumption behavior of the resident is identified. In such of case we further study how the existing clusters have evolved in the newly built clustering model in order to investigate whether this discrepancy is associated with alternations in the resident's sleep habits.

2.1 k-medoids

Three partitioning algorithms are commonly used for data analysis to divide the data objects into k disjoint clusters [MacQueen, 1967]: k-means, k-medians, and k-medoids clustering. The three partitioning methods differ in how the cluster center is defined. In k-means clustering, the cluster center is defined as the mean data vector averaged over all objects in the cluster. In k-medians, the median is calculated for each dimension in the data vector to create the centroid. Finally, in k-medoids clustering, which is a robust version of the k-means, the cluster center is defined as the object with the smallest sum of distances to all other objects in the cluster, i.e., the most centrally located point in a given cluster.

2.2 Dynamic Time Warping

The DTW alignment algorithm aims at aligning two sequences of feature vectors by warping the time axis iteratively until an optimal match (according to a suitable metrics) between the two sequences is found [Sakoe and Chiba, 1978].

Let us formally explain how DTW works. Consider two matrices $A=[a_1,...,a_n]$ and $B=[b_1,...,b_m]$ with a_i $(i=1,...,n)$ and b_j $(j=1,...,m)$ column vectors of the same dimension. The two vector sequences $[a_1,...,a_n]$ and $[b_1,...,b_m]$ can be aligned against each other by arranging them on the sides of a grid, *e.g.* one on the top and the other on the left hand side. A distance measure, comparing the corresponding elements of the two sequences, can then be placed inside each cell. To find the best match or alignment between these two sequences one needs to find a path through the grid $P=(1,1),...,(i_s,j_s),...,(n,m)$, $(1\leq i_s\leq n$ and $1\leq j_s\leq m)$, which minimizes the total distance between A and B. Thus, the DTW distance between A and B can be defined as

$$\mathrm{dtw}(A,B)=\frac{1}{n+m}\min_P\left(\sum_{s=1}^{k}d(i_s,j_s)\right).$$

3 Initial Evaluation and Results

3.1 Data

In this study we use electricity consumption data collected from a single randomly selected anonymous household that has been collected with a 1-minute interval. We divide the data into two sets, one to represent the historical data that is used to model the ordinary electricity consumption behavior of the household (data set A) and the other one to represent the new portion of data used to track changes in the existing household's behavior model (data set B). Data set A contains 242 (70%) of the initial profiles and B the remaining 104 profiles (30%). For both sets we aggregate the electricity consumption data from 1-minute measures to 1-hour measures. Previous studies have indicated that using a too fine granularity makes it harder to identify patterns of consumption [Nordahl *et al.*, 2017].

In order to study and evaluate the effectiveness of our approach, we introduce noise to data set B by altering a number of daily profiles. We insert 4 different amounts of noise to this data set, 5% (B_5), 10% (B_{10}), 15% (B_{15}), and 20% (B_{20}), and we retain one without any noise (B_0). The noise insertion is performed by

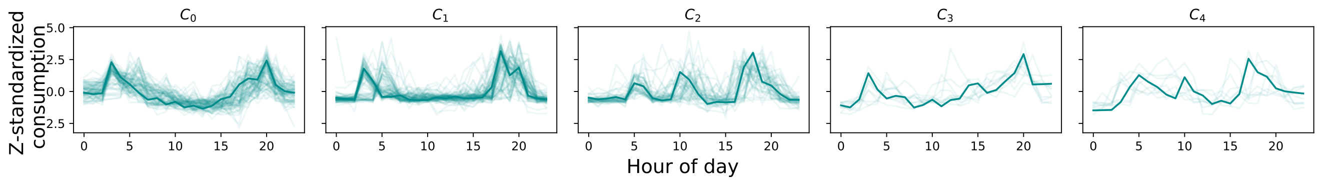

Figure 2: Cluster profiles of the clustering solution created by using data set A

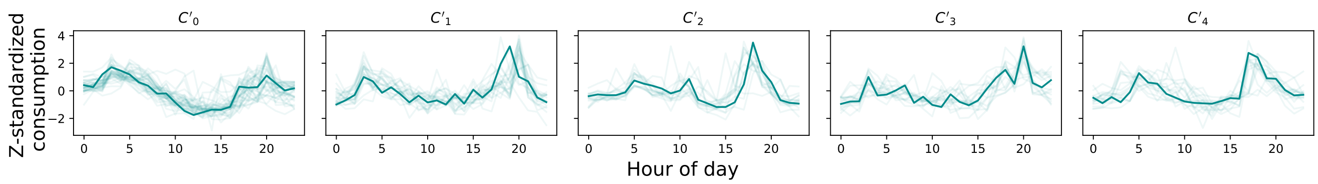

Figure 3: Cluster profiles of the clustering solution created by using data set B_0

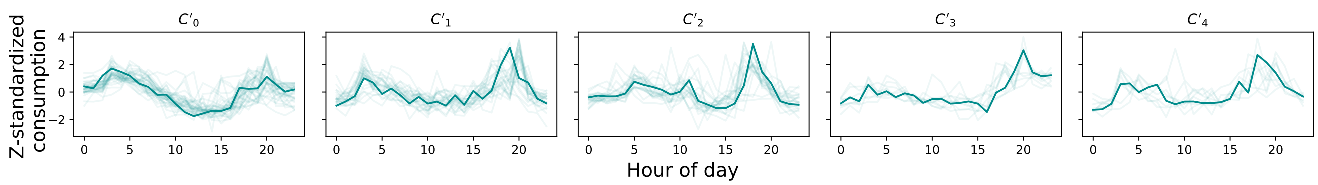

Figure 4: Cluster profiles of the clustering solution created by using data set B_5

shifting one day x hours forward, where x is either (i) 12 hours or (ii) a random number of hours between 1 and 6. We choose these two different scenarios to mimic either (i) a complete change of the individual's sleeping pattern (12 hour shift) or (ii) a slighter deviation of the normal sleeping routine.

Then for all data sets we perform a z-standardization on all individual profiles, as we are more interested in the actual shapes during the studied period, i.e. how they consume, and not how much they consume. We argue that a normal behavior could look similar during the summer and winter, but the amplitudes may differ.

3.2 Experimental Setup

Initially, we apply k-medoids clustering on data set A in order to build the ordinary electricity consumption behavior model of the household. The optimal number of clusters (k) is determined by initially studying and evaluating the used data by three different cluster validation indices: Silhouette Index [Rousseeuw, 1987], Connectivity [Handl *et al.*, 2005], and Average Intra-Cluster distance [Baya and Granitto, 2013]. In our experiments data set B represents the new portion of data collected over a predefined time period. In order to receive an insomnia diagnosis, patients must experience trouble falling or staying asleep for a period of one month or longer [Roth, 2007]. Therefore in our experiments the time axis is partitioned in monthly or longer intervals, i.e. the data set B covers a period longer than month.

We conduct five different experiments by considering five different new data sets: B_0, B_5, B_{10}, B_{15}, and B_{20} (see Section 3.1). Each data set B_i, for $i = 0, 5, 10, 15, 20$, is clustered by using k-medoids initialized by the medoids of the clustering solution generated on data set A.

	B_0	B_5	B_{10}	B_{15}	B_{20}
A	0.0907	0.0955	0.0955	0.0889	0.0921

Table 1: The DTW distances between the existing cluster solution produced on data set A and the cluster solutions generated by the new data sets B_0, B_5, B_{10}, B_{15} and B_{20}.

3.3 Results and Discussion

Table 1 lists the DTW distances between the existing cluster solution produced on data set A and the cluster solutions generated by the new data sets B_0, B_5, B_{10}, B_{15} and B_{20}, respectively. One can notice the highest distances are recorded for data sets B_5, B_{10} and B_{20}. Therefore, for the rest of this section we focus on studying the results produced by these three data sets and comparing them to the ones generated by data set B_0. The latter can be considered as a baseline for the new portion of data, since it has not been injected with any noise.

Table 2 and Table 3 present the DTW distances calculated between the cluster medoids of the existing clusters and the new clusters generated on data sets B_0 and B_{20}, respectively. It is interesting to observe and trace the evolution of clusters C_1 and C_4 in data set B_{20} compared to data set B_0. For example, C_1 gets closer to C'_2, while C_4 moves towards clusters C'_2 and C'_3. These observations are also supported by the heatmaps plotted in Figure 5 and Figure 6, respectively. One can observe that the respective cells in the heatmap of B_{20} have changed their colour in comparison with the heatmap of the baseline data set B_0. In addition, as it can be seen in Table 4, cluster C'_4 which is a transition of cluster C_4 in the newly generated clustering solution gets shrunk in all three studied

	C'_0	C'_1	C'_2	C'_3	C'_4
C_0	0.198	0.239	0.268	0.220	0.225
C_1	0.354	0.174	0.241	0.268	0.187
C_2	0.472	0.219	0.182	0.317	0.216
C_3	0.334	0.221	0.408	0.170	0.299
C_4	0.377	0.237	0.259	0.318	0.182

Table 2: DTW distances between the cluster medoids of the existing clusters (C_0,...,C_4) and the new clusters (C'_0,...,C'_4) generated on data set B_0.

	C'_0	C'_1	C'_2	C'_3	C'_4
C_0	0.198	0.239	0.335	0.252	0.249
C_1	0.354	0.174	0.182	0.256	0.218
C_2	0.472	0.219	0.164	0.304	0.293
C_3	0.334	0.221	0.364	0.199	0.228
C_4	0.377	0.237	0.213	0.281	0.186

Table 3: DTW distances between the cluster medoids of the existing clusters (C_0,...,C_4) and the new clusters (C'_0,...,C'_4) generated on data set B_{20}.

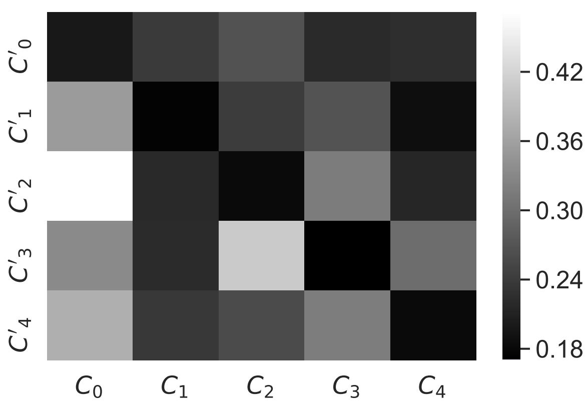

Figure 5: Heatmap for distances between the existing clusters and the clusters generated by B_0.

data sets (B_5, B_{10} and B_{20}), while logically cluster C'_2 gets large.

We study further the above discussed observations in Figures 2, 3 and 4, which depict cluster profiles of the clustering solutions generated on data sets A, B_0, B_5, respectively. C_0 and C_1 are the biggest clusters and represent the electricity consumption behavior more typical for working days with clearly recognized morning and evening consumption peaks. Cluster C_2 is also comparatively big and has an additional consumption peak in the middle of the day, i.e. it models behavior more typical for the weekends. Clusters C_3 and C_4 are smaller and do not represent so clearly recognizable behavior. Now let us trace how these five clusters are evolved in the new data sets B_0 and B_5, respectively. As we can see in Table 4 clusters C'_0 and C'_1 are also the biggest ones in both data sets (B_0 and B_5) and have very similar shapes to those of C_0 and C_1 (see Figures 3 and 4). However, cluster C'_2 does not have a similar profile to that of its origin cluster C_2. It models behavior that shows electricity consumption activity too early in the morning and in addition, it expands in data set B_5.

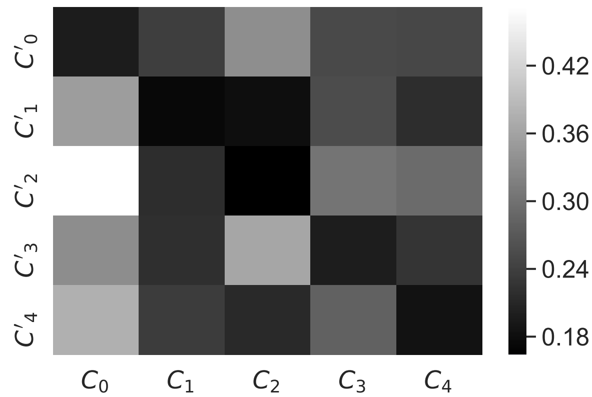

Figure 6: Heatmap for distances between the existing clusters and the clusters generated by B_{20}.

	C'_0	C'_1	C'_2	C'_3	C'_4
B_0	36	21	15	16	16
B_5	36	25	22	10	9
B_{10}	36	23	21	13	9
B_{15}	34	20	15	14	19
B_{20}	37	23	22	10	10

Table 4: Distribution of daily profiles to different clusters for all B data sets.

The transformation of cluster C_4 in data sets B_0 and B_5 also has an effect on its shape. We can observe that C'_4 does not have an additional consumption peak in the middle of the day like C_4.

4 Conclusions and Future Work

In this paper, we have proposed an approach for monitoring household electricity consumption behavior and mining changes that can be used to recognize shifts in daily routines and sleeping patterns of the resident. The proposed approach uses clustering techniques for modelling and analyzing household's electricity consumption behavior. It has been initially evaluated on electricity consumption data collected from a single randomly selected anonymous household. The experimental results have shown that the approach is robust to recognizing alternations in the resident daily routines by monitoring and analyzing their electricity consumption behavior.

Our future plans are to pursue further evaluation of the proposed approach by involving healthcare experts and using data from other sources, e.g. ambient sensors, for its validation on richer data and real scenarios. In a long-term perspective, we are interested in applying the developed approach for recognizing early signs of dementia by monitoring and tracing changes in household electricity consumption behavior in addition to other data sources.

Acknowledgements

This work is part of the research project "Scalable resource-efficient systems for big data analytics" funded by the Knowledge Foundation (grant: 20140032) in Sweden.

References

[Baya and Granitto, 2013] Ariel E Baya and Pablo M Granitto. How many clusters: A validation index for arbitrary-shaped clusters. *IEEE/ACM Trans. on Comput. Biol. and Bioinformatics*, 10(2):401–414, 2013.

[Brzecka *et al.*, 2018] Anna Brzecka, Jerzy Leszek, Ghulam Md Ashraf, Maria Ejma, Marco F. Ávila Rodriguez, Nagendra S. Yarla, Vadim V. Tarasov, Vladimir N. Chubarev, Anna N. Samsonova, George E. Barreto, and Gjumrakch Aliev. Sleep disorders associated with alzheimer's disease: A perspective. *Frontiers in Neuroscience*, 12:330, 2018.

[Buet, 2017] Solene Buet. *Analysis of sleep-related symptoms of Parkinson's patients based on a system of ambient sensors*. Master Thesis, ETHZ, Department of Information Technology and Electrical Engineering, Zurich, Switzerland, 2017.

[Cirpiani *et al.*, 2015] Gabriele Cirpiani, Claudio Lucetti, Sabrina Danti, and Angelo Nuti. Sleep disturbances and dementia. *Psychogeriatrics,Japanese Psychogeriatric Society*, 15:65–74, 2015.

[de Almondes *et al.*, 2016] Katie Moraes de Almondes, Monica Vieira Costa, Leandro Fernandes Malloy-Diniz, and Breno Satler Diniz. Insomnia and risk of dementia in older adults: Systematic review and meta-analysis. *Journal of Psychiatric Research*, 77:109–115, 2016.

[Handl *et al.*, 2005] Julia Handl, Joshua Knowles, and Douglas B Kell. Computational cluster validation in post-genomic data analysis. *Bioinformatics*, 21(15):3201–3212, 2005.

[MacQueen, 1967] James MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.

[Nordahl *et al.*, 2017] Christian Nordahl, Marie Persson, and Håkan Grahn. Detection of residents' abnormal behaviour by analysing energy consumption of individual households. In *1st Workshop on Aging, Rehabilitation, and Independent Assisted Living (ARIAL) 2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 729–738. IEEE, 2017.

[Nordahl *et al.*, 2018] Christian Nordahl, Veselka Boeva, Håkan Grahn, and Marie Netz. Organizing, visualizing and understanding households electricity consumption data through clustering analysis. In *Proceedings of the 2nd ARIAL Workshop, IJCAI 2018*, pages 29–32, Stockholm, Sweden, July 2018.

[Nordahl *et al.*, 2019] Christian Nordahl, Veselka Boeva, Håkan Grahn, and Marie Netz. Profiling of household residents' electricity consumption behavior using clustering analysis. In *Proceedings of International Conference on Computational Science*, Faro, Algarve, Portugal, June 2019. Springer.

[Ping-Ru *et al.*, 2015] Ko Ping-Ru, Julie A Kientz, Eun Kyoung Choe, Matthew Kay, Carol A Landis, and Nathaniel F Watson. Consumer sleep technologies: A review of the landscape. *Journal of Clinical Sleep Medicine*, 11(12):1455–1461, 2015.

[Roth, 2007] Thomas Roth. Insomnia: definition, prevalence, etiology, and consequences. *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, 3(5 Suppl):S7, 2007.

[Rousseeuw, 1987] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[Sakoe and Chiba, 1978] Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1):43–49, 1978.

[Shelgikar *et al.*, 2016] Anita Valanju Shelgikar, Patricia F Anderson, and Marc R Stephens. Sleep tracking, wearable technology, and opportunities for research and clinical care. *Chest*, 150(3):732–743, 2016.