Opening the Personal Gate between Technology and Health Care
M. Giacomini et al. (Eds.)

doi:10.3233/SHTI260724

Bridging Data and Behavior in Homecare: Personalized Routine Modelling and Anomaly Interpretation

RAJA OMMAN ZAFAR[a,1] and YVES RYBARCZYK
[a] *Dalarna University, Sweden*
ORCiD ID: Raja Omman Zafar https://orcid.org/0000-0002-7223-7977

Abstract. Understanding how older adults organize their daily lives is crucial for developing person-centered homecare systems. This study proposes an interpretable framework for modeling daily life and detecting behavioral anomalies using data from 18 CASAS smart homes. The dataset contains several weeks of continuous sensor recordings from residents living independently. Daily activity patterns were analyzed in 15-minute intervals using principal component analysis (PCA) to identify key temporal patterns shared by the population. For each resident, a personal baseline routine was defined as the median of their daily activity profiles over a 14-day baseline period, and deviations from this baseline were compared with global deviations derived from the PCA model. The results revealed explainable behavioral differences among residents and highlighted three lifestyle archetypes like active bimodal, stable routine, and early resting. By linking the difference scores to contextual activities such as sleep, hygiene, and computer use, the framework provides relevant explanations for daily irregular behaviors.

Keywords. Behavioral modeling, homecare monitoring, anomaly interpretation, ambient intelligence, human activity analysis

1. Introduction

The increasing integration of ambient intelligence and the Internet of Things (IoT) is transforming how everyday activities are monitored and analyzed in smart home environments. Modern sensor systems continuously record data on movement, environmental conditions, and appliance usage, creating a detailed picture of how people live and interact within their homes [1][2]. These data streams offer unique opportunities for objectively understanding human behavior and provide valuable insights into lifestyle, health, and well-being. In person-centered homecare settings, these systems can promote autonomy, detect early signs of behavioral changes, and enable proactive interventions [3]. However, human routines are complex, context-dependent, and inherently variable, which makes it challenging to extract meaningful behavioral patterns from raw sensor data. Managing this complexity requires modeling approaches that can transform data into interpretable and actionable knowledge [4].

A wide range of methods have been proposed to model human behavior in intelligent environments. Supervised learning methods, such as probabilistic models and deep neural networks, have achieved high accuracy in activity recognition tasks by learning from labeled datasets [5]. These models can identify activities such as sleeping, eating,

[1] Corresponding Author: Raja Omman Zafar, roz@du.se

or cooking, but require extensive annotations and often fail to generalize to new users or environments [3]. In contrast, unsupervised and semi-supervised methods, including clustering, autoencoders, and anomaly detection by reconstruction, can reveal patterns directly from unlabeled data [6]. While such methods alleviate the need for manual labeling, they typically lack interpretability and provide limited insight into the behavioral meaning of detected anomalies.

Despite these advances, several research gaps remain. Most existing research treats each individual as an isolated unit, focusing either on identifying activities within a household or on building broad group behavior models [7]. This separation limits our ability to capture both collective patterns and individual differences within a unified framework. Furthermore, many supervised and deep learning-based anomaly detection system rely on labeled data, reducing their interpretability and practical suitability for real-world homecare deployment [8]. To address these challenges, this paper proposes an explainable framework for modeling generalized-personalized daily activities and explaining anomalies in homecare. Leveraging multi-resident data from 18 Centre for Advanced Studies in Adaptive Systems (CASAS) smart homes, the framework combines group-level pattern discovery through PCA with personalized baseline modeling for each resident. Correlating deviations with contextual activity tags further reveals the human meaning behind detected anomalies, while clustering in PCA space uncovers lifestyle archetypes across residents.

2. Methodology

This study utilized the publicly available CASAS smart home dataset, a recognized benchmark for behavior modeling and activity recognition [9]. The study analyzed data from 18 residents living independently in sensor-equipped homes, covering 30 to 45 consecutive days for each participant. The CASAS system records data from multiple environmental sensors, including motion, contact, temperature, light, and appliance sensors, capturing a variety of daily activities such as cooking, sleeping, and computer use. All data was anonymized and made publicly available by the dataset authors to ensure ethical compliance and privacy.

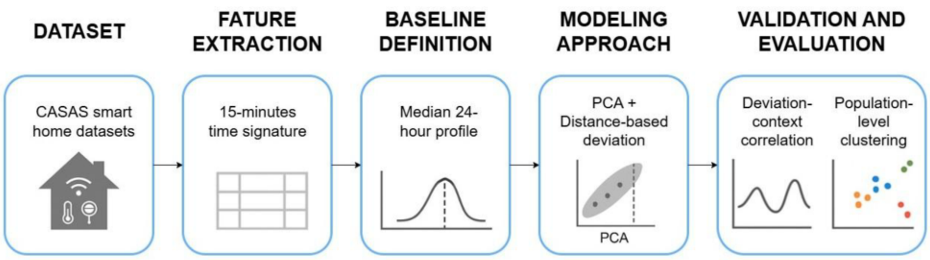

Figure 1. Framework for personalized routine modeling and anomaly interpretation

To convert raw event data into analyzable behavioral patterns, sensor activations were aggregated into fixed 15-minute intervals, resulting in 96-time segments representing a full day. Days with missing or incomplete data were excluded, resulting in approximately 630 daily features. To ensure comparability, all characteristics were z-score standardized to reduce bias associated with differences in sensor density or housing layout. For each resident, a 14-day reference period was defined to determine their daily habits. The median of this period represented the individual's most stable behavior pattern, and subsequent days were compared to this individual reference value and the population model to assess bias.

This modeling framework integrates generalization and personalization through a hybrid design. A PCA model is trained on aggregated reference data to capture global behavioral patterns, such as circadian rhythms and activity peaks [10]. Individual deviations are calculated as the Euclidean distance from each individual's reference profile, while global deviations are derived from the PCA reconstruction error. A composite deviation index averages these two metrics to identify atypical days for both individuals and groups. To assess interpretability, we correlated bias with indicators of contextual activities such as sleep and computer work, confirming that high bias corresponds to significant behavioral differences. Finally, to identify lifestyle patterns at the population level, we projected residents' average daily activity levels into a principal component analysis (PCA) space and clustered them using the k-means clustering algorithm. The number of clusters (k = 3) was determined based on the elbow and silhouette coefficient criteria to balance simplicity and interpretability. This analysis revealed three consistent lifestyle prototypes like stable routine, sedentary/early-resting, and active bimodal.

3. Result

The proposed framework effectively captures both general and individual behavioral variations among 18 CASAS residents. Figure 2 summarizes the key findings, showing user-level irregularity, population-level behavioral patterns, clustering in the PCA space, and lifestyle archetypes. As shown in Figure 2a, residents exhibit significant differences in their average deviation levels. Some individuals (e.g., users 7 and 8) exhibit high irregularity, indicating significant variability in their daily behavior, while others (e.g., users 2, 13, and 16) display highly consistent daily behavior. These differences demonstrate the model's ability to quantify individual behavioral stability and detect lifestyle differences across participants.

At the group level, the PCA model extracted interpretable behavioral components that represent the most prominent temporal rhythms in daily life (Figure 2b). The first three principal components explained approximately 16% of the total variance of the 96-time interval variables (15-minute intervals), which is consistent with prior studies using fine-grained activity data, where the variance is distributed across many independent temporal features [10]. These components concisely and meaningfully characterize collective human behavior, confirming that PCA can serve as an interpretable foundation for understanding temporal activity patterns in smart homes.

By projecting users into PCA space and applying unsupervised clustering, three distinct lifestyle archetypes emerged like active bimodal, stable routine, and early resting (Figure 2c-d). Residents of the active bimodal group exhibited two significant activity peaks in the morning and evening, indicating a regular work-life schedule. The stable routine group maintained a balanced and predictable daily activity, while the early resting group exhibited lower activity levels and earlier rest times, suggesting decreased mobility or a preference for early bedtimes. These archetypes emerged naturally from unsupervised learning, highlighting the framework's ability to reveal interpretable, data-driven lifestyle patterns that may facilitate early detection of behavioral changes in older adults.

4. Discussion

The results demonstrate that combining group-level modeling with individual-level baseline analysis provides a practical and interpretable approach for understanding behavioral patterns in homecare environments. The high correspondence between overall and individual deviations confirms the framework's ability to capture both common and specific resident dynamics. Cases of individual deviations do not reflect group-level variation, highlighting the importance of individual context in defining normal behavior. Unlike supervised approaches, the proposed PCA-based framework emphasizes transparency and low deployment burden by operating entirely on unlabeled data. Despite its simplicity, this approach provides a structured way to describe behavioral variation using interpretable metrics rather than complex predictions. It can more clearly explain how residents' habits change over time, providing a clear baseline for future, more advanced modeling efforts.

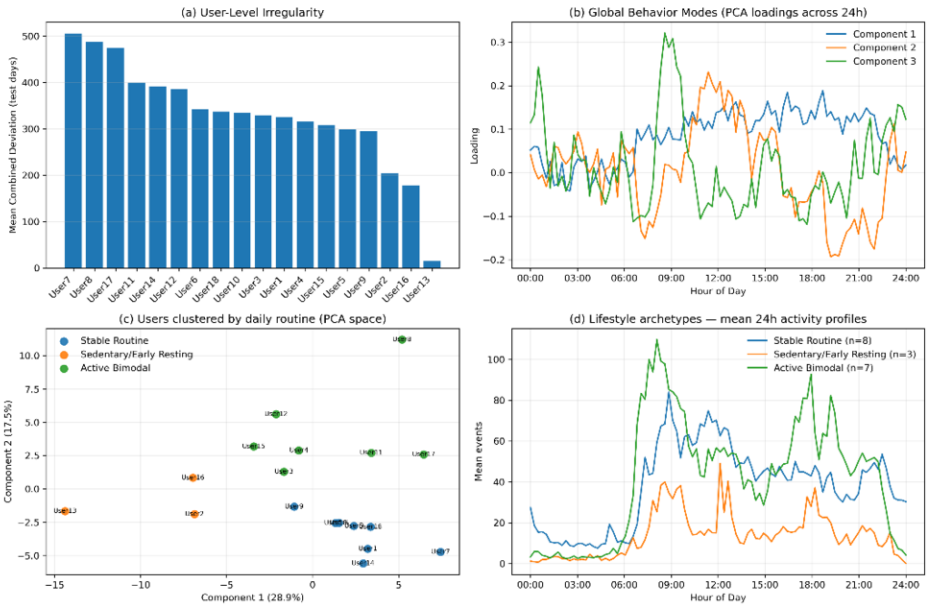

Figure 2. Behavioral analysis results (a) showing resident variability, (b) principal components, (c) clustering in PCA space, and (d) three lifestyle archetypes.

Grouping residents into lifestyle archetypes demonstrates the framework's ability to derive interpretable insights from unlabeled data. The identified archetypes like active, stable routine, and early resting reflect the natural diversity of daily human rhythms and align with established patterns of behavioral variation. This interpretability is crucial in homecare settings, as gradual shifts in activity or rest cycles can signal early signs of cognitive or physical decline. The primary beneficiaries include homecare professional and care coordinators those require transparent indicators of routine stability rather than automated diagnosis. While promising, this research is limited by the relatively small data size, high data homogeneity, and the use of linear PCA, which can miss nonlinear dependencies in complex behaviors. Future research should extend this framework to larger multi-household datasets and incorporate nonlinear or hybrid deep learning techniques to improve sensitivity without compromising interpretability [3]. Integrating this approach into a real-time digital twin system could further support adaptive home care by continuously learning daily habits and identifying emerging deviations that indicate health-related changes [11].

5. Conclusion

In conclusion, this study proposes an explanatory framework combining population modeling and personalized baseline analysis to interpret daily behavior in homecare environments. By combining global patterns based on PCA with individual deviation measures, this approach captures both shared and unique aspects of daily human behavior while maintaining interpretability. The identification of lifestyle archetypes and meaningful behavioral deviations demonstrates its potential for transparent and actionable monitoring. By prioritizing interpretability and minimal labeling requirements, this work can support the development of trustworthy decision-support tools for personalized and preventative homecare.

References

[1] Udupa P, Yellampalli SS. Smart home for elder care using wireless sensor. Circuit World. 2018 May;44(2):69-77, doi: 10.1108/CW-12-2017-0072.

[2] Zafar RO, Latif I. Exploring Advanced Deep Learning Architectures for Older Adults Activity Recognition. In: Miesenberger K, Peňáz P, Kobayashi M, editors. Computers Helping People with Special Needs: 19th International Conference, ICCHP 2024; 2024 Jul 8–12; Linz, Austria. Cham (CH): Springer Nature Switzerland; c2024. p. 320–327, doi: 10.1007/978-3-031-62849-8_38.

[3] Zafar RO, Zafar F. Real-time activity and fall detection using transformer-based deep learning models for elderly care applications. BMJ Health Care Inform. 2025 Sep;32(1), doi: 10.1136/bmjhci-2025-101439.

[4] Cejudo A, Beristain A, Almeida A, Rebescher K, Martín C, Macía I. Smart home-assisted anomaly detection system for older adults: a deep learning approach with a comprehensive set of daily activities. Med Biol Eng Comput. 2025 Jun;63(6):1821–1835, doi: 10.1007/s11517-025-03308-y.

[5] Bouchabou D, Nguyen SM, Lohr C, LeDuc B, Kanellos I. A Survey of Human Activity Recognition in Smart Homes Based on IoT Sensors Algorithms: Taxonomies, Challenges, and Opportunities with Deep Learning. Sensors. 2021 Sep;21(18), doi: 10.3390/s21186037.

[6] Ige AO, Mohd Noor MH. A survey on unsupervised learning for wearable sensor-based activity recognition. Appl Soft Comput. 2022 Sep;127:109363, doi: 10.1016/j.asoc.2022.109363.

[7] Vani MS, Sudhakar RV, Mahendar A, Ledalla S, Radha M, Sunitha M. Personalized health monitoring using explainable AI: bridging trust in predictive healthcare. Sci Rep. 2025 Aug;15(1):31892, doi: 10.1038/s41598-025-15867-z.

[8] Alaghbari KA, Md Saad MH, Hussain A, Alam MR. Activities recognition, anomaly detection and next activity prediction based on neural networks in smart homes. IEEE Access. 2022;10:28219–28232, doi: 10.1109/ACCESS.2022.3157726.

[9] Cook DJ, Crandall AS, Thomas BL, Krishnan NC. CASAS: A smart home in a box. Computer (Long Beach Calif). 2013 Jul;46(7), doi: 10.1109/MC.2012.328.

[10] Wood MD, Simmatis LER, Jacobson JA, Dukelow SP, Boyd JG, Scott HH. Principal components analysis using data collected from healthy individuals on two robotic assessment platforms yields similar behavioral patterns. Front Hum Neurosci. 2021 May;15:652201, doi: 10.3389/fnhum.2021.652201.

[11] Zafar RO, Rybarczyk Y, Borg J. A systematic review of digital twin technology for home care.ACM Trans Comput Healthcare. 2024 Oct;5(4):1–24, doi: 10.1145/3681797.