Unobtrusive Activity Detection of Daily Living Activities Using Binary Sensors

Cecilia Hunter
Jamie Horton
c.r.hunter@nmsu.edu
jhorton@nmsu.edu
New Mexico State University
Las Cruces, New Mexico, USA

Keywords

elder monitoring, disability monitoring, binary sensors, health and safety, machine learning, non-invasive

1 Problem statement

Increasingly, people live alone. Work takes people away from their families, long- or short-term. The world's population is aging, and increasingly doing so at home, either from choice or necessity. Young people move frequently, and digital work and social opportunities allow people to go for days or longer without seeing other people in person. With this increase, people are also noticing that this comes with unique safety risks. Life alone means that if someone is sick or injured, they may be unable to call for help. In China, an app called "Are You Dead" requires checking in on a regular basis. If a check is missed, it will text the user's preferred emergency contact[4]. Similar apps are available internationally, including some which connect to wearable smart devices or phone sensors. However, for many people, a smartphone app is not an ideal solution. Disability, technological savvy, memory issues, or other factors can make checking in with an app unreliable or impossible. Other health concerns may require a greater level of monitoring, such as notifications for higher risk of injury or illness.

Sensor data, such as infrared devices, cameras, door switches, and wearables have been considered as possible solutions to this problem. Wearables not only have to be both charged and worn reliably, but add an additional layer of possible data insecurity [5]. Additionally, many forms of safety monitoring come with privacy concerns. GPS monitoring may be too inaccurate for determining if someone is safe in their home, while providing too much information about their location outside of the home. Camera technology, whether infrared or visual spectrum, may also be a privacy concern [3], especially because cameras may indicate specific activities, rather than merely presence. For example, while it is useful to know if someone has been in the bathroom an unusually long time, it may not be appropriate for a particular monitor to know exactly what activities took place. Increasingly, though, non-invasive sensors such as binary door and motion sensors, as well as temperature sensors are available and affordable, often through companies already prepared with monitoring services, such as home security providers. Previous studies have sought to detect significant anomalies in behavior which are indicative of health or safety concerns warranting human intervention, such as injury or illness, performing known behaviors in unusual spaces, performing known behaviors at significantly unusual times, or performing known behaviors with unusual durations[1–3, 5]. Being able to detect anomalies quickly and inform an appropriate emergency contact or service could provide significant improvements in safety of people living alone.

However, studies in this area have primarily sought to determine activity patterns and anomalies for persons living alone. The ability to identify patterns of activity and anomalies in homes with multiple bodies present is especially crucial to developing practical monitoring systems for both aging adults and other persons with disabilities for which monitoring could be valuable. This would allow monitoring to account for normal human activities such as living with other people, having company visit the home, and having pets, and receiving intermittent caregiving services. Additionally, being able to detect the activities of caregivers could assist not only with monitoring the health and wellness of a person receiving them, but could assist in monitoring safety and well-being for both the client and caregiver, ensuring caregiver presence as scheduled, safety monitoring to prevent abuse in either direction, and appropriate task completion.

2 Methods

We frame our monitoring task under two complementary machine learning formulations. 1. Supervised Multi-Class Time-Series Classification When activity labels are available, (e.g. annotated smart home datasets such as CASA), we are able to treat the problem of activity identification as a supervised multi-class time series classification task. By then identifying similarity between an "average" day, and the observed, we can look for anomalies in both a single day, and in patterns of increasing dissimilarity from the norm. 2. When labels are unavailable or sparse, we treat the problem as semi-supervised anomaly detection. In this setting, we assume that most observed behavior reflects normal daily activity, and anomalie correspond to statistically rare or structurally unusual patterns. This provides for anomalies to include prolonged inactivity, unusual room transitions, environmental readings outside learned ranges, or behavior or room occupation occuring at atypical times of day.

We have identified two datasets for our initial exploration, each using different kinds of sensors to determine human presence. Both have been collected in real residential environments, inhabited by a single elderly individual, and were designed to support research in elder monitoring, anomaly detection, and behavior modeling in home environments. Because both reflect real-world, non-invasive sensing modalities, they align well with our research motivation of privacy-preserving home monitoring.

First, the CASAS Smart home dataset for home HH126. This provides the data for a single home with one occupant living independently in a retirement community, from August 2013 through July 2014, using two kinds of sensors: passive infrared sensors detecting movement through an on/off status, and magnetic sensors detecting open or closed status. Sensors are identified by their location in the home. An additional dataset is available for this home, which includes labels for activities completed by an external annotator for August 1, 2013 through September 6, 2013. Finally, a floor plan is available for this data, to provide context for the sensor data. The annotated dataset consists of 111,801 timestamped records, while the non-annotated dataset has 813,109 records. These datasets are publicly available through the Center for Advanced Studies in Adaptive Systems (CASAS). Additionally, CASAS has made many similar datasets of sensor data available, both with a single adult in a retirement community, but also including single people, couples, and families in homes. In all, this level of sensor data is available for 189 homes, and was generated between 2007 and 2024. Many of these sets also have annotated partial datasets and floor plans, allowing us to explore whether models developed can generalize well to similar data.

The Environmental and Gas Sensor Time Series consists of three datasets. The first contains timestamped room-level sensor events, including binary activations for six rooms in the home, and has 20,478 rows of data, covering a span from November 1, 2019, through January 24, 2020. This data has a median inter-event interval of approximately 30 seconds, and is appropriate for modeling activities based on time window aggregation and room transition analysis. Once a predictable pattern of activities is determined, it is then possible to detect anomalies based on similarity measures to an average day, for each day of the week. The second dataset within this series provides continuous environmental measurements including temperature, humidity, carbon dioxide proxies, carbon monoxide values, and metal oxide gas sensors. It has approximately 416,100 rows, in the same time span as the first in the series, and a sampling interval of approximately 20 seconds. The third set is the secondary gas dataset, which provides information from January 25, 2020 through February 13, 2020 with the same structure as Dataset 2, but while the house was unoccupied, in order to provide baseline readings which can be used for comparison.

3 Feature Engineering

Raw sensor logs are converted into structured machine learning inputs using window-based aggregation. For binary sensor data, time is segmented into fixed-length windows, sensor activations per room are counted, time since last activation is computed, room transition frequency is encoded, and contextual features such as the hour of the day, day of the week, and weekend indicators are included. For environmental sensors, a rolling mean and standard deviation are computed, the first differences (slopes) are calculated, features are normalized, and we can optionally compute windowed variance or energy metrics.

4 Candidate Models and Rationale

We propose evaluating the following models:

- Logistic regression, which provides a highly interpretable baseline and benchmark
- Random Forest, which captures non-linear relationships and handles noisy features
- Gradient Boosting, which can model complex feature interactions efficiently
- One-Class SVM, to provide a boundary around normal behavior
- Hidden Markov Models (HMM), which capture sequential dependencies and state transitions
- LSTM networks, which learn long-range temporal dependencies
- Isolation forests, which provide effective unsupervised anomaly detection.

To prevent data leakage and preserve temporal realism, we use time-aware splits, training on earlier data, and testing on later data. Metrics include accuracy, macro-F1 score, precision, recall, confusion matrices, ROC-AUC for anomaly detection, and PR-AUC when anomalies are rare. Hyperparameter tuning will be conducted using cross-validation within the training set. In supervised settings, anomalies may include rare activity classes or misclassified windows. In unsupervised settings, anomalies are defined as statistically significant deviations from learned normal patterns.

5 Initial Data Exploration

Initial analysis has included event counts per room, hour-of-day activity patterns, histograms of environmental variables, and correlation analysis among gas sensors. Basic pre-processing and exploratory scripts have been implemented to load datasets, parse timestamps, compute summary statistics, evaluate sampling intervals, and generate visualizations. This establishes the foundation for later modeling stages, including classification and anomaly detection.

6 Results

We hope the resulting templates and documentation will help the readers to write submissions for ACM journals and proceedings.

7 Significance

This document is important for anybody wanting to comply with the requirements of ACM publishing.

8 Bibliography

References

[1] H. Chojer, P.T.B.S. Branco, F.G. Martins, M.C.M. Alvim-Ferraz, and S.I.V. Sousa. 2022. Can data reliability of low-cost sensor devices for indoor air particulate matter monitoring be improved? – An approach using machine learning. *Atmospheric Environment* 286 (Oct. 2022), 119251. doi:10.1016/j.atmosenv.2022.119251

[2] Daejin Kim, Hongyi Bian, Carl K. Chang, Linag Dong, and Jennifer Margrett. 2022. In-Home Monitoring Technology for Aging in Place: Scoping Review. *Interactive Journal of Medical Research* 11, 2 (Sept. 2022), e39005. doi:10.2196/39005

[3] Daniel Marín, Joshua Llano-Viles, Zouhair Haddi, Alexandre Perera-Lluna, and Jordi Fonollosa. 2023. Home monitoring for older singles: A gas sensor array system. *Sensors and Actuators B: Chemical* 393 (Oct. 2023), 134036. doi:10.1016/j.snb.2023.134036

[4] Stephen McDonell. 2026. *Are You Dead?: The viral Chinese app for young people living alone.* Retrieved February 2, 2026 from https://www.bbc.com/news/articles/c3381r5nnn6o

[5] T. H. Tan, L. Badarch, W. X. Zeng, M. Gochoo, F. S. Alnajjar, and J. W. Hsieh. 2023. Binary Sensors-Based Pirvacy-Preserved Activity of Elderly Living Alone Using an RNN. *Sensors* 21 (Oct. 2023), 14248220. Issue 16. doi:10.1016/j.snb.2023.134036