
There may be differences between this version and the published version. You are advised to consult the published version if you wish to cite from it.

https://eprints.gla.ac.uk/298669/

Deposited on 19 May 2023

Enlighten – Research publications by members of the University of Glasgow http://eprints.gla.ac.uk
Room-Level Activity Classification from Contextual Electricity Usage Data in a Residential Home

Muhammad Farooq†*, Mahmoud A. Shawky‡*, Aisha Fatima†*, Ahsen Tahir†, Muhammad Z. Khan†,
Hasan T. Abbas‡, Muhammad Imran‡, Qammer H. Abbasi†, and Ahmad Taha†

†James Watt School of Engineering, University of Glasgow, Glasgow, G12 8QQ, United Kingdom
‡Electrical Engineering, University of Engineering and Technology, 54 890, Pakistan

Email: {m.farooq.1, m.shawky.1}@research.gla.ac.uk, 2798454F@student.gla.ac.uk, ahsan@uet.edu.pk,
m.khan.6@research.gla.ac.uk, {Hasan.Abbas, Muhammad.Imran, Qammer.Abbasi, Ahmad.Taha}@glasgow.ac.uk

Abstract—Human activity recognition is challenging without compromising users’ privacy and burdening them with wearable devices, cameras, mobile applications, etc. As the smart energy meter usage trend is increasing worldwide, it can be used as a non-invasive activity monitoring methodology without violating users’ privacy and requiring an additional installation cost where smart energy meters are already in use. In addition, household energy consumption patterns, including active and reactive power, current intensity, and energy usage, are mainly determined by the individual’s needs, lifestyle, and time context, which can offer important information about the household’s daily activities. Using energy, one can get information about ongoing activities and the types of loads. This paper uses the Random Forest, Decision Tree, K-Nearest Neighbour, and Support Vector Machines-based machine-learning algorithms for activity recognition and load classification. In this context, an open smart energy meter dataset is employed to recognize the activity patterns of house occupants. Based on the four types of machine-learning algorithms, the load classification analyses show an accurate recognition probability of up to 99%.


I. INTRODUCTION

Recently, activity monitoring technology has emerged as an efficient method to assist decision-making and reduce the burden on caretakers to detect emergencies concerning elderly people [1], [2]. The current state-of-the-art for activity monitoring typically depends on smartphone applications and internet of things (IoT) devices, including cameras, specialized sensors, smart plugs, etc to monitor individual health in real-time [3], [4]. In [5], vision-based activity recognition systems are used to increase detection accuracy. Although the existing works offer accurate detections, they may violate users’ privacy. Further to the significant cost and burden incurred by installing and distributing these IoT devices. For example, homeowners are reluctant to install cameras in their homes for privacy reasons; other occupants are uncomfortable wearing and charging wearable devices.

In this challenging scenario, the energy consumption-based activity monitoring method has been introduced as a privacy-preserving method. This method avoids the aggravation of purchasing, wearing, and installing various sensors and devices. The only information required for activity detection is energy consumption data from an existent smart energy meter and a smart device that can compute a baseload, compare continuously incoming data with the baseload, and predict activities. Electricity usage data can accurately identify operational devices and their locations in rooms accurately when compared to baseload and device load descriptors. Previous works of activity detection approach based on power disaggregation data, commonly known as non-intrusive load monitoring (NILM) is presented in [6]. Contrary to the existing work, our approach focuses on activity detection utilizing electricity usage data. The proposed work employs energy consumption data as a privacy-preserving methodology compared to vision-based approaches. The following summarises the contribution of this work.

- Using machine-learning algorithms, we propose an efficient activity-level classification throughout the year based on fine-grained electricity usage data from a sub-metering system.
- We highlighted the concept behind using high-resolution electricity usage data to recognize different activities in a day.

The remainder of this paper is structured as follows. Section II provides a brief review of related works. Section III introduces the proposed machine learning-based classification approach. While Section IV discusses the classification results. Finally, Section V concludes the paper.

II. RELATED WORKS

Within an advancement in IoT, smart homes are introduced to facilitate users by providing several home services, i.e., health care monitoring, privacy, etc. In smart homes, many actuators and sensors are used to control daily life equipment and they are linked together with communication protocols. Now, human activity recognition techniques are being developed by researchers to monitor and analyze behaviour of people using sensors and actuators. Generally, human activity recognition can be categorized into vision-based and sensor-based methods [7], as shown in Fig. 1. The former systems use cameras and computer-based vision techniques; however, younger adults (till age 35) are more concerned with privacy and are not happy with sharing and tracking of their activities [8]. While sensor-based systems are further divided
into three categories, i.e., wearables, sensors on objects and ambient sensors. Use of sensor-based smart homes is a feasible solution [9]. However, human activity recognition is a challenging because everyone has his own way of living a life. Several algorithms are proposed and implemented for human activity recognition in smart homes, including pattern recognition, feature extraction algorithms, etc. Pattern recognition techniques are further divided into two categories, i.e., data-driven and knowledge-driven techniques. In-depth knowledge about the domain is required in a knowledge-driven approach; however, domain experts are needed to make new rules which may replace previous rules. This limitation is later solved by a data-driven approach, i.e., supervised and unsupervised both methods are included in this approach for human activity recognition using energy data. These approaches do not require prior knowledge; further, they can handle uncertainties in data, i.e., noise, incomplete data, etc. However, much clear, correctly labeled data and more computational time is required for a data-driven approach [10]. To summarize, use of deep learning and machine-learning models in a data-driven approach for human activity recognition using energy data is a viable solution to train more complex models and they do not invade people’s privacy [4]. In this regard, this paper uses the Random Forest, Decision Tree, K-Nearest Neighbour, and Support Vector Machines-based machine-learning algorithms for activity recognition and load classification by considering the privacy of individuals.

III. THE PROPOSED APPROACH

In this section, we introduce the methodology adopted to perform energy consumption-based activity classification using machine-learning.

A. Activity recognition

In energy consumption-based activity recognition, the amount of energy consumed during a certain period of time during the day is correlated with the activity of the home occupants. By learning the daily consumed energy of a house, known by the baseload, one can predict the type of activity and investigate any anomalies. The type of activity in the house can be discerned by comparing electricity usage data with the baseload, operated load, and location in the house. In this paper, we evoked the household electric power consumption data in [11]. This dataset is generated from a sub-metering system that monitors three different rooms in the house and reports the energy consumption in Watt-hour (Wh), as shown in Fig. 2. The data has been presented for each day of the year in the form of $[dd/mm/yyyy]$ at a specific time $[hh:mm:ss]$. The estimated consumed energy from each room is recorded every minute (sampling rate) for the whole year. Each room has a list of electric devices, each of which has a unique power rate (PR) represented as follows:

- **Room (1)**: The kitchen contents are a dishwasher (PR: $\sim$ 1200 : 2400 W), and an oven (PR: $\sim$ 2000 : 5000 W).
- **Room (2)**: The laundry room contents are a washing machine (PR: $\sim$ 2000 : 2500 W), a tumble dryer (PR: $\sim$ 2000 : 2500 W), a refrigerator (PR: $\sim$ 300 : 800 W), and a light (PR: $\sim$ 6 : 60 W).
- **Room (3)**: The contents are an electric water heater (PR: $\sim$ 1375 : 4245 W), and an air-conditioner (PR: 3500 W).

As a proof of concept, we aggregated the energy consumed by each room every minute within a time interval of half an hour for the whole day. Fig. 3 presents the aggregated consumed energy for a randomly selected day from each season. It can be seen from Fig. 3 that the activities of the house occupants in all rooms are low at the night hours from [23 : 00 : 00] to [06 : 00 : 00]. Contrary to the morning hours, activities are observed in Room (1) ”Kitchen” at breakfast and lunch times. Meanwhile, we can observe other activities in Room (2) due to high energy consumption from the tumble dryer and washing machine (e.g., see Fig. 3(a) between [14 : 00 : 00] and [16 : 00 : 00]) which can indicate that the residents are doing their laundry. Furthermore, as can be observed in Fig. 3, the behaviour of Room (3)’s sub-meter, in all sub-figures (a) to (d), indicates that the air-conditioning unit and/or the water heater were actively used during the afternoon hours and occasionally at night, despite the season.
B. Machine-learning-based load classification approach

In a two-phase process, we train and test the machine-learning kernel using the energy consumption data related to different activities in each room at different time slots \((t)\). The training and testing phases are further elaborated on in this sub-section.

As a first step, we split the energy consumed every half hour of a specific activity \((x)\) into five categories, represented as follows:

- **Class (1)**: \(200 \text{ Wh} > x \geq 100 \text{ Wh}\).
- **Class (2)**: \(300 \text{ Wh} > x \geq 200 \text{ Wh}\).
- **Class (3)**: \(400 \text{ Wh} > x \geq 300 \text{ Wh}\).
- **Class (4)**: \(500 \text{ Wh} > x \geq 400 \text{ Wh}\).
- **Class (5)**: \(600 \text{ Wh} > x \geq 500 \text{ Wh}\).

Note that, we did not consider very low consumption data \((x < 100 \text{ Wh/hr})\) for the entire year, as it indicates no holding activity. For the training purposes of different machine-learning algorithms, we use 80\% of the data in the same form presented in Fig. 3 as a training, while the rest of the data, 20\%, is used for testing purposes. The classifier output categorizes each room’s daily activities into five classes.

In this study, we investigated this approach with different types of machine-learning algorithms, including Decision Tree (DT), K-Nearest Neighbour (KNN), Random Forest (RF), and Support Vector Machine (SVM).

The classification algorithms were trained and tested on each individual room, separately, to classify and identify the activity happening within a specific period of time.

IV. RESULTS AND DISCUSSION

Four machine-learning classification algorithms are applied to the electricity data to recognize daily activities in three rooms. The classification results are presented in Fig. 4. In this study, the number of instantaneous occurrences of a specific activity \(x\) in the whole year is denoted by \(|x|\).

As can be seen from Fig. 4, \(|x|_{\text{Room(1)}} \sim 17\) times/year for class (5) at \(t = [12 : 00 : 00]\), while this value at \(t = [19 : 30 : 00]\) is about \(\sim 33\) times/year. In other words, we observe high energy consumption activities in Room (1) at this time of the day. Similarly for Room (2) and (3), \(|x|_{\text{Room(2)}}\) for class (5) is higher than that for other classes at \([22 : 00 : 00] > t > [12 : 00 : 00]\). In Room (3), \(|x|_{\text{Room(3)}} \sim 225\) times/year at \(t = [09 : 00 : 00]\), which means high energy consumption activity in the range of class (5) occurred around 225 times/year with probability of daily occurrence equals \(\frac{220}{365} \times 100 = 60\%\). Any holding activities out of the range of the probability of occurrence at a certain time slot can be considered anomalies.

In Table I, you can see an example of the classification report produced by the SVM-based algorithm. As can be seen, the precision, recall, and F1-score are about 0.99, indicating precise classification. In Table II, the accuracies of classification for the four types of machine-learning algorithms.
are given. The four types of classifiers give high detection accuracy of about $\sim 0.99$, proving the practicality of the proposed approach to support efficient activity recognition. As a future work, we can build an additional layer of a machine-learning model that can classify specific activities across the entire house based on different classifications from the rooms’ models, and apply it to our own dataset.

**V. Conclusion**

This paper proposes an efficient way to classify activities based on energy consumption data generated from a residential building’s sub-metering system. Four types of machine-learning algorithms are used to classify the daily activities of house occupants for each room at different times. Through the machine-learning classification algorithms, which had an accuracy of $\sim 99\%$, associated with information on the appliances monitored by each sub-meter, we were able to identify the number of occurrences of using each room throughout the day and across the whole year. With more fine-grained data, this method can be further exploited to develop systems that can give indications of abnormal activities that can be identified from deviations in the usual consumption patterns, which can potentially be used to predict critical events. In the future, we will explore ways to design an additional layer of

**REFERENCES**


