Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 9 Issue 37 (December 2024) PP. 133-146
DOI: 10.35631/JISTM.937011

JOURNAL OF INFORMATION SYSTEM AND TECHNOLOGY MANAGEMENT (JISTM)
www.jistm.com

TIME SERIES PATTERNS RECOGNITION WITH DYNAMIC DATA REDUCTION IN NILM APPLIANCE IDENTIFICATION

Kwan-Hua Sim[1*], Saad Tariq[2], Kwan-Yong Sim[3]

[1] School of Information and Communication Technologies, Swinburne University of Technology Sarawak, Malaysia
Email: khsim@swinburne.edu.my
[2] School of Information and Communication Technologies, Swinburne University of Technology Sarawak, Malaysia
Email: stariq@swinburne.edu.my
[3] School of Information and Communication Technologies, Swinburne University of Technology Sarawak, Malaysia
Email: ksim@swinburne.edu.my
[*] Corresponding Author

Article Info:

Article history:
Received date: 24.10.2024
Revised date: 29.11.2024
Accepted date: 08.12.2024
Published date: 22.12.2024

To cite this document:

Sim, K. H., Tariq, S., & Sim, K. Y. (2024). Time Series Patterns Recognition With Dynamic Data Reduction In NILM Appliance Identification. *Journal of Information System and Technology Management, 9* (37), 133-146.

DOI: 10.35631/JISTM.937011

Abstract:

The convergence of advancements of Internet of Things (IoT) capabilities, coupled with the accessibility of inexpensive and user-friendly sensors, has propelled the emergence of various new domains, and one such domain is Non-Intrusive Load Monitoring (NILM). A pivotal aspect of these technological advancements is the identification of appliances through the analysis of disaggregated power consumption signatures. The length of these signatures is contingent upon the frequency of data collection, where higher frequencies correspond to lengthier time series. To address this, we introduce a novel dynamic time series data reduction methodology, tailored to efficiently extract regions of interest from extended time series data. Subsequently, the efficacy of appliance classification using these extracted sub-ranges is assessed through the utilization of Matrix Profile techniques. The experimental validation is conducted utilizing the Plug-Load Appliance Identification Dataset, thus offering a concrete empirical basis for our approach's evaluation and verification. The proposed approach has successfully improved the overall accuracy of identifying the appliances in PLAID dataset with a significant margin as compared to the baseline approach.

Keywords:

"Classification Methods", "Data Analytics", "Information And Knowledge", "Time-Series Model", "Time-Series Patterns".

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 9 Issue 37 (December 2024) PP. 133-146
DOI: 10.35631/JISTM.937011

Introduction

The confluence of technological progress in Internet of Things (IoT) capabilities, coupled with the proliferation of cost-effective and user-friendly sensor technology, has engendered a substantial upsurge in the volume of time series data generated across diverse domains. This fast-growing data landscape has subsequently catalyzed advancements within correlated disciplines such as time series data mining, similarity search methodologies, and subsequence matching techniques (Sarhan, 2023). Notably, one of the domains that closely relates to IoT time series pattern recognition is known as Non-Intrusive Load Monitoring (NILM) (Hernández et al., 2019).

Non-Intrusive Load Monitoring (NILM), also known as Non-Intrusive Appliance Load Monitoring (NIALM) or disaggregated energy monitoring, is a technology and methodology that aims to determine the individual energy consumption patterns of specific appliances within a household or industrial setting without requiring direct physical access to the appliances themselves (Dash, 2022). By comparing the observed aggregate signal with a database of known appliance signatures or using sophisticated algorithms, NILM systems can estimate the operation of various appliances and provide insights into their energy consumption (Hafiz et al., 2021).

The ultimate goal of NILM is to enable users to better understand their energy usage patterns, identify energy-wasting appliances, and make informed decisions about energy conservation and management. It has applications in energy efficiency optimization, demand-side management, and building automation, among others.

In the context of advancements in IoT and sensor technology, Non-Intrusive Load Monitoring (NILM) is intricately linked with energy disaggregation and load identification. These applications conventionally bifurcate their operational objectives into two fundamental tasks, specifically denoted as energy disaggregation and load identification (Hafiz et al., 2021). Recent investigations have increasingly centered on pattern matching methodologies to facilitate the discernment of loads on the basis of their characteristic signature patterns (Dash, 2022).

Essentially, one primary challenge is the scalability and complexity of handling vast time-series data generated by IoT devices, which can lead to significant computational overhead and inefficiencies, especially with dynamic datasets. Additionally, accuracy in appliance identification can be hampered by overlapping signal patterns and transient variations, making reliable energy disaggregation a challenge. The lack of standardization in IoT protocols and methods adds to interoperability challenges, hindering seamless integration across devices and systems (Rana et al, 2021).

Over the course of scholarly exploration, substantial endeavors have been devoted to investigating a range of methodologies within the domain of pattern matching and recognition, which hold relevance in the foundational principles of Non-Intrusive Load Monitoring (NILM). In particular, approaches such as real and reactive powers, the geometrical features of V-I plots, current, instantaneous power, and harmonics have been exploited to identifying the signals (Tezde et al., 2022).

Journal of Information System and Technology Management
EISSN: 0128-1666
Volume 9 Issue 37 (December 2024) PP. 133-146
DOI: 10.35631/JISTM.937011

Notably, the development of a basic algorithm predicated on nested iterative loops is inevitably confronted by the escalation computational complexity as the size of the energy consumption dataset expands. This escalation engenders a critical resource-intensive overhead, encompassing both temporal and computational dimensions, which causes the practical implementation of such an approach infeasible for the voluminous and dynamic energy consumption data sets typically encountered in NILM applications (Himeur et al., 2022).

Thus, this study aims to develop novel methodologies that streamline the process of identifying appliances and their energy usage patterns. It involves the exploration and application of dynamic data reduction techniques, which are designed to efficiently distill pertinent time series patterns from voluminous time series data collected from various electrical appliances. Additionally, a state-of-the-arts time series pattern search algorithm, Matrix Profile (MP) will be implemented on the time-series data produced by the proposed data reduction technique.

Since its inception, the Matrix Profile has found extensive application across a myriad of temporal data analysis and mining challenges. The pervasive adoption of this data structure finds its rationale in the manifold benefits it affords, notably its inherent domain-agnostic nature, exceptional computational efficiency, dependency on a singular parameter (namely, the sliding window size) (Yeh, et al. 2016).

The implementation of the proposed approach entailed the selection of the Plug-Load Appliance Identification Dataset (PLAID) (Medico et al. 2020). as dataset for all the experiment. The rationale underlying this selection is multi-fold. Notably, the dataset offers a markedly elevated sampling frequency of 30 kHz during data acquisition, making it conducive to optimal appliance identification endeavors.

Each recorded measurement in PLAID encapsulates a dual set of parameters, comprising the instantaneous current and voltage. Encompassing 1876 documented instances of individually metered appliances covering 16 distinct appliance categories, the dataset's comprehensive composition amplifies its significance in the empirical validation of the proposed approach. Five instances are then randomly selected from each class to be the sample (or the unknown query) which will be classified.

Subsequently, a dimension reduction technique is utilized on the extracted sample of subsequences to retain only the significant segments. These extracted subsequences are subsequently matched against the collection of load signatures in the existing database by employing the Matrix Profile (MP) method. This process generates a distinct similarity value for each appliance class, serving as the foundation for the classification of an unknown sample.

As such, the scope of this study revolves around leveraging advancements on the Internet of Things (IoT) and sensor technology to address challenges in Non-Intrusive Load Monitoring (NILM). It focuses on optimizing the identification of individual appliance energy consumption patterns through innovative methodologies like dynamic data reduction and Matrix Profile (MP) computation. The goal is to enhance accuracy and efficiency in appliance classification within NILM applications by minimizing computational complexity while handling extensive time-series data.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 9 Issue 37 (December 2024) PP. 133-146
DOI: 10.35631/JISTM.937011

Literature Review

Non-Intrusive Load Monitoring (NILM) has emerged as a pivotal technology for enhancing energy efficiency and promoting sustainable energy consumption practices. Central to NILM is the capability to disaggregate aggregated energy consumption data into individual appliance-level. Achieving accurate appliance identification from such data streams hinges on effective time series pattern recognition techniques.

The primary challenge encountered in time series pattern recognition resides in the computational intricacy inherent to the comprehensive search process. Among the notable methodologies that have gained prominence is the Mueen and Keogh (MK) algorithm, a renowned approach that has garnered recognition for its substantial enhancement of time complexity when contrasted with the conventional brute-force technique. Notably, the MK algorithm retains the capacity to deliver precise time series pattern matches, signifying its robust efficacy. This algorithm operates at a time complexity of O(nR), wherein 'R' denotes the count of user-defined reference points, contributing to its efficient pattern recognition attributes (Deppe, et al., 2022).

In the realm of recent advancements, substantive strides have been taken in the arena of time series pattern matching, predominantly attributed to the advent of the Matrix Profile framework. This mechanism substantially facilitates pragmatic time series pattern matching across extensive datasets, which capitalizes on innovative computational techniques to tangibly reduce the time complexity to O(log(n)) for the precise identification of time series patterns within a given length (Deppe, et al., 2022).

The formula to compute the Matrix Profile for a time series "T" of length "n" using a subsequence length "m" is given by:

$$MP[i] = \sqrt{\sum_{j=1}^{m}(T[i+j-1] - T[j])^2} \tag{1}$$

where:

- MP[i] is the Matrix Profile value at index "i"
- T[i+j−1] is the data point of the time series at index i+j−1i+j−1
- T[j] is the data point of the query subsequence at index j
- m is the subsequence length
- The summation iterates over all points within the subsequence

In essence, the Matrix Profile at each index "i" reflects the similarity or distance be-tween the subsequence of the time series starting at index "i" and its nearest neighbor within the entire time series. This concept aids in identifying motifs or repeated pat-terns within the time series data (Zymbler, at el., 2021).

The introduction of the matrix profile represents a notable advancement in the field of time series data analysis. In essence, the matrix profile annotates a time series by providing information about the position and distance to the nearest neighbor for each subsequence. Before the advent of the matrix profile, the maximum dataset length for precise time series mining was one million data points. This limitation was surpassed with the introduction of the Matrix Profile, which expanded the capacity by 100 times. Notably, Zhu et al. (2016) sought to explore time series subsequences with a length of 60,000 in the bioinformatics domain.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 9 Issue 37 (December 2024) PP. 133-146
DOI: 10.35631/JISTM.937011
The Matrix Profile framework necessitates only one single parameter from the user, specifically the desired length of the time series subsequence. This marks a significant advancement over numerous preceding methodologies, which often demanded the specification of multiple parameters. However, it is imperative to underscore that while Matrix Profile presents a streamlined approach, it lacks a comprehensive solution for time series patterns with varying lengths (Linardi et al., 2018).

In the context of NILM implementation, the operation involves many different stages, but load identification is the most critical stage. However, as energy consumption patterns grow increasingly complex, the accurate identification of individual appliances from aggregated energy signals becomes a significant challenge.

The predefined time series patterns as well as the instantaneously observed patterns are frequently matched in template matching-based load detection mechanism. Basu et al. direct template matching is the outcome of direct measuring of the distances between the observed time series subsequence pattern and the template pattern (Basu, et al. 2016). Bouhouras et al. (2019) suggested reducing the difference between the target time series pattern and the combination time series pattern.

Traditional load identification techniques relied on heuristic approaches, often leveraging distinctive signatures like power and voltage levels, harmonic patterns, and transient behaviors. While these techniques provided valuable insights into energy consumption, they faced limitations when confronted with the complexities of real-world scenarios. Overlapping appliance signatures and transient load variations posed challenges to these methods, leading to inaccuracies in load identification (Tezde at el, 2022).

In response to these challenges, the application of advanced signal processing techniques has emerged as a critical avenue for enhancing load identification within the context of NILM. Waveform analysis techniques such as Fourier Transform and Wavelet Transform have been pivotal in extracting spectral and frequency-domain features from energy consumption data (Kang et al., 2020).

Meanwhile, time-domain techniques like Correlation and Cross-Correlation have been instrumental in detecting synchrony between aggregated energy signals and known appliance load signatures. By measuring the similarity between the aggregated signal and reference appliance signatures, these methods facilitate the assignment of energy consumption to specific appliances (Jaradat et al. 2020), (Wójcik et al., 2021). However, their efficacy can be hampered by the presence of noise and overlapping appliance activations.

Additionally, data-driven techniques such as Singular Value Decomposition (SVD) and Principal Component Analysis (PCA) have gained traction in the domain of load identification. These techniques transform the high-dimensional energy consumption data into a reduced-dimensional space, effectively highlighting the significant variance within the data. This transformation enhances the separation between distinct appliance consumption patterns, thereby improving the accuracy of load identification, especially in scenarios with varying load conditions (Wójcik et al., 2021), (Moradzadeh et al. 2020).

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 9 Issue 37 (December 2024) PP. 133-146
DOI: 10.35631/JISTM.937011
Hidden Markov Models (HMMs) are another widely used framework for time-series analysis, offering a probabilistic approach to modeling sequential data. In the context of NILM, HMMs are particularly effective due to their ability to model the hidden states of appliances and the observable aggregate energy signals (Malik et al., 2021). An HMM represents the system as a set of hidden states connected by transition probabilities, with each state emitting observable signals based on a defined probability distribution. This capability makes HMMs highly suitable for disaggregating energy consumption data, as they can capture the temporal dependencies and patterns inherent in appliance usage (Sankara, 2014).

One of the key advantages of HMMs in NILM is their ability to handle noise and uncertainty in the aggregated energy signals, which are common in real-world data. By leveraging probabilistic state transitions, HMMs can infer the most likely sequence of appliance states over time, even when signals overlap or exhibit transient variations. For multiple appliances that operate simultaneously, HMMs can disentangle their contributions by learning the characteristic emission probabilities of each appliance. This is especially useful for appliances with distinct power consumption patterns, as the model can distinguish between overlapping signals based on state likelihoods (Miquey et al., 2021).

Despite their strengths, HMMs have limitations that impact their scalability and effectiveness in complex NILM scenarios. The computational complexity increases significantly with the number of appliances and states, making it challenging to apply HMMs to large-scale datasets with numerous devices (Wu et al., 2021a). Moreover, traditional HMMs assume stationary transition probabilities and emissions, which may not align with the dynamic nature of real-world energy consumption. Extensions like factorial HMMs and variable-order HMMs have been proposed to address these limitations, offering more flexibility in modeling complex interactions and capturing longer-term dependencies (Wu et al., 2021b).

Despite the notable progress in time series pattern discovery for NILM, challenges persist. Real-world energy consumption data are inherently variable due to user behavior, appliance degradation, and changing load conditions. This variability necessitates adaptive algorithms that can account for dynamic patterns and transitions. Moreover, the scalability of pattern recognition techniques remains a concern, as they must efficiently handle extensive datasets with numerous appliances and high-frequency measurements.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 9 Issue 37 (December 2024) PP. 133-146
DOI: 10.35631/JISTM.937011

Methodology

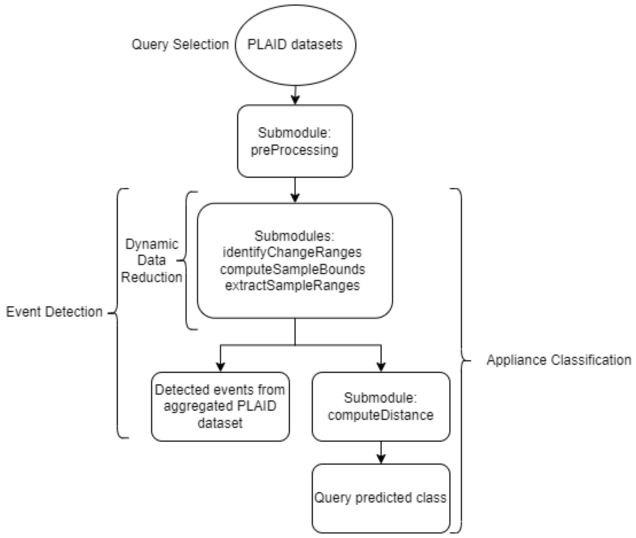

Figure 1: High-level Overall Architecture Diagram

Figure 1 illustrates the high-level architecture diagram that begins with query selection, where a query is chosen from either the submetered or aggregated dataset obtained from PLAID. This selected query is then passed through the preprocessing submodule, which performs various operations to compute the power consumption signature and extract other relevant information.

The preprocessed query is subsequently passed through three additional submodules, collectively known as Dynamic Data Reduction. These submodules work together to analyse and reduce the data in order to capture the essential event information.

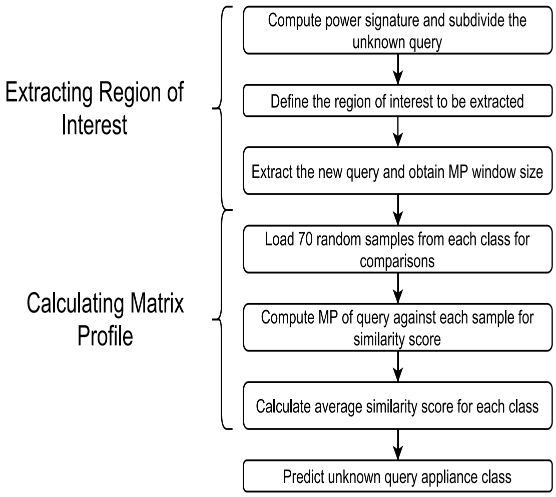

Figure 2: Dynamic Data Reduction Approach

The experimental approach comprises two main components. The first component involves the extraction of regions of interest from the selected sample or query of the raw NILM dataset through the deployment of a novel dynamic dimension reduction technique. The subsequent

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 9 Issue 37 (December 2024) PP. 133-146
DOI: 10.35631/JISTM.937011
phase entails the computation of Matrix Profiles (MPs) for these extracted ranges, subsequently yielding similarity scores. Notably, the possibility of extracting multiple regions of interest exists, depending upon the behavior exhibited by the appliances.

As outlined in Figure. 2, the dynamic data reduction process begins by breaking down the selected sample into smaller segments, a step taken to precisely identify the significant areas of interest. The size of these smaller segment is determined by a percentage parameter specified by the user. This parameter is influenced by the sampling frequency employed during data collection. Importantly, a higher sampling frequency corresponds to narrower sections, creating an inverse relationship between the frequency and the width of the sections.

After dividing the entire sample into smaller segments, it is followed by calculating the disparities in the standard deviation between each segment and its adjacent counterpart. These deviations are then averaged across all the segments. Subsequently, the individual deviations of each segment are compared to this calculated average. This methodology hinges on a fundamental notion – the fluctuation in power consumption that arises when an appliance is powered on or undergoes state changes. This leads to greater deviation in segments with such changes, as opposed to those with consistent energy consumption patterns.

As a result, segments displaying higher deviations than the calculated average are deemed of interest. Importantly, multiple segments often meet this criterion. If these segments are contiguous, they are combined into a continuous pattern. This consolidated pattern replaces the initial query, now shorter and more focused. Appliances that can be classified as on/off devices typically yield a single region of interest. However, appliances categorized as finite state machines or those with continuous variations often result in multiple regions of interest. In these cases, all the extracted subsequences are treated as equally important and are employed in the subsequent phase of the process.

A noteworthy insight arises from the examination of the plots of NILM datasets, indicating that the utilization of the complete time series is unnecessary. This discovery affords the opportunity to truncate the extent of the query, thereby ensuring its consistency relative to the sample undergoing scrutiny. This stipulation is indeed a prerequisite for the proper computation of Matrix Profiles (MPs). Furthermore, it is pertinent to specify a window size parameter, but unless the user possesses substantial domain expertise and can precisely ascertain the optimal size, this this is frequently a trial-and-error process. Remarkably, the incorporation of a dimension reduction step effectively mitigates this limitation as well.

As outlined in Figure. 2, the second component, calculating Matrix Profile is implemented on the extracted subsequences, an iterative process is undertaken to calculate the Matrix Profiles (MPs) of the subsequence query using various time series from the dataset. Given that the subsequence query represents our dynamically extracted regions of interest, the length of the newly formed query is crucial as it functions as our designated window size. Consequently, there is no necessity to ascertain this parameter for each iteration.

For instances where only a single region is extracted from the samples, a computation is performed for each pair, and the outcomes are stored in a dictionary data structure. In this structure, the values corresponding to each key, where the key represents time series identifiers, denote their computed Matrix Profiles (MPs) after the calculations. Utilizing these MPs, the

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 9 Issue 37 (December 2024) PP. 133-146
DOI: 10.35631/JISTM.937011
algorithm identifies the smallest distance within each pair. Subsequently, the algorithm computes the average value for each class.

Once all the requisite MPs are computed, the algorithm derives an averaged similarity score for each class. This process encapsulates the overall approach's methodology for generating similarity scores and facilitating class-wise comparisons.

On the other hand, In cases where multiple regions are extracted from the samples, Matrix Profiles (MPs) are calculated for all these regions in relation to the respective time series. Subsequently, the smallest distances are determined from each of these computed MPs. Among these distances, the minimum value is selected as the ultimate similarity score. All other distances, along with the corresponding MPs and their associated regions, are disregarded and excluded from further consideration.

Following this step, the pertinent information is stored in a manner consistent with the previously mentioned approach. This method of computation and selection ensures that the final similarity score is derived from the most relevant and representative distance, thereby contributing to the overall accuracy of the similarity comparison process.

The method begins with the random selection of a sample from the database, followed by the computation of its power consumption signature. Subsequently, a dimension reduction technique is applied to this newly generated time series. This orchestrated process encapsulates the essence of our dynamic data reduction strategy, which is poised to enhance the efficiency and accuracy of subsequent analyses and load identification endeavors.

Experiments
A series of experiments were carried out to empirically showcase the efficacy of the proposed approach in the context of load identification within the framework of Non-Intrusive Load Monitoring (NILM). These experiments and analyses, this section aims to substantiate the methodology's potential to accurately distinguish appliances based on their energy consumption patterns. By leveraging proposed approach in dynamic data reduction, Matrix Profile computation, and similarity scoring, the experiments delve into the intricate process of load identification across a diverse set of appliances.

The PLAID dataset encompasses a total of 16 distinct appliance classes, although there exists a notable discrepancy in the number of instances associated with each class. For instance, examining the classes "Blender," "Vacuum," and "Fan," we find they possess 2, 83, and 220 instances, respectively. To warrant a robust and balanced evaluation when classifying unknown queries, an important measure is undertaken. Specifically, classes with fewer than 50 instances, such as "Blender," "Coffee Maker," "Hair Iron," "Soldering Iron," and "Water Kettle," are omitted from consideration.

This pragmatic approach aims to ensure that the evaluation process remains statistically significant and equitable across all appliance classes. Among the remaining classes, any instance selected at random holds the potential to serve as the query for analysis. In each instance, there are two essential parameters: current consumption and voltage consumption. While the current parameter offers visually obvious signatures, the product of these two

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 9 Issue 37 (December 2024) PP. 133-146
DOI: 10.35631/JISTM.937011

parameters, known as the apparent power, holds the promise of yielding enhanced accuracy (Angelis, et al., 2022).

After the random query has been selected, the respective significant pattern of the power is identified. Essentially, the length of the extracted subsequence will subsequently become the query for MP in the next stage, which is the window size parameter in MP. Following the execution of this sub-routine, the algorithm proceeds to the next step, where it calculates Matrix Profiles (MPs).

Following the random selection of instances from each class, the iterative phase of Matrix Profile (MP) computation commences. Once these computations are completed, the next step involves calculating the average of the smallest distances obtained from the 70 instances within each class. This consolidation of distances serves as a representative measure. Ultimately, the class with the lowest average similarity score emerges as the predicted class for the unknown sample.

Findings & Discussion

To ensure equitable comparisons, a total of five random instances were selected from each class. It's important to emphasize that no prior information was gathered regarding these instances prior to their selection. This approach fosters an objective and unbiased evaluation of the experimental methodology within the context of Non-Intrusive Load Monitoring (NILM).

To establish a baseline, the initial step involved computing the average Euclidean Distance (ED) for every selected instance within each class. Between the length of the sample and the length of the said instance, the shorter time series was kept as the query and the calculation was done throughout the longer time series. In this manner, two Euclidean Distance (ED) scores were acquired for every pair of time series, and the smaller of the two ED values was retained.

Subsequently, this process was iterated for all 70 instances within each class, and the average for each class was calculated. The smallest value recorded was the predicted class of the unknown query. The classification accuracy for this approach is depicted in Figure 3, revealing that none of the classes reached an accuracy of 100%. The class exhibiting the highest accuracy in classification was "CFL," achieving an 80% accuracy rate, indicating the highest number of random instances were correctly identified. Following closely, the "Microwave" category secured the second-highest accuracy at 60%. However, the accuracy rates for the "Hairdryer" and "Vacuum" classes were 40%, while classes such as "Fan," "AC," and "Heater" achieved a lower accuracy of only 20%. Regrettably. not a single instance was accurately classified from the remaining categories.

To evaluate the effectiveness of the proposed approach, the entire experiment was replicated with the inclusion of the dynamic data reduction technique applied to the queries. The results of this modified experiment are illustrated in Figure 4. Notably, a significant enhancement in classification accuracy is evident across all classes. Instances from four out of the 11 classes achieved 100% accuracy in classification, while classes such as "ILB," "AC," and "Laptop" obtained an accuracy of 80%. The classes "Hairdryer," "Fridge," and "Fan" achieved accuracies of 60%, 40%, and 20%, respectively. This decline in accuracy is likely attributed to the inherent characteristics of the signals, as further discussed later. However, not a single instance of "Washing Machine" was accurately identified in both experiments.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 9 Issue 37 (December 2024) PP. 133-146
DOI: 10.35631/JISTM.937011

Another evaluative metric is the overall accuracy of Euclidean Distance (ED) comparisons versus the proposed approach, depicted in Figure 5. When the least computed ED was employed for classifying unknown queries, only 14 out of the 55 randomly chosen samples were correctly classified, resulting in a 25.45% accuracy. However, with the implementation of dynamic data reduction, a total of 38 out of 55 samples were accurately classified, yielding a significantly improved accuracy of 69.09%.

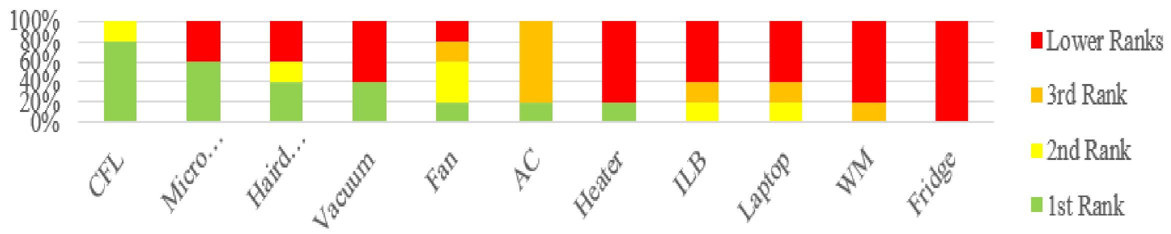

Figure 3: Classification Accuracy of Unknown Sample (Baseline)

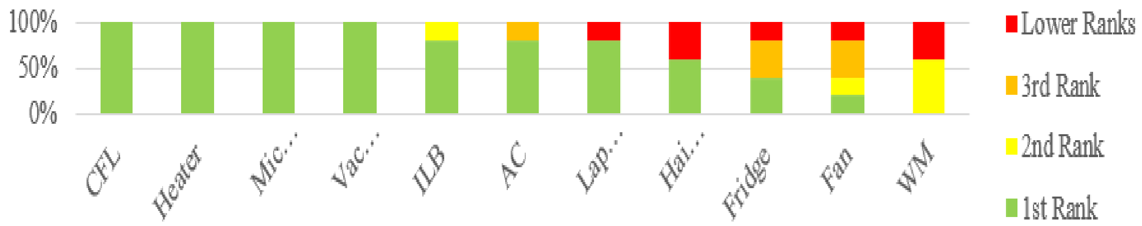

Figure 4: Classification Accuracy of Unknown Sample (Proposed Approach)

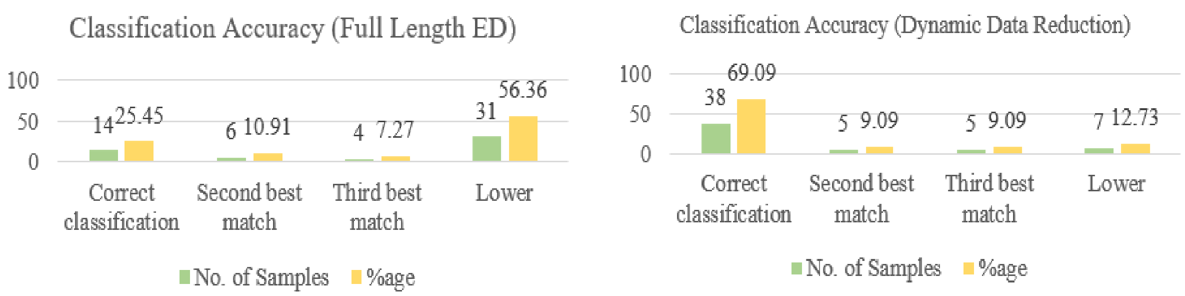

Figure 5: Overall Classification Accuracy Between Baseline (right) Versus the Proposed Approach (left)

An observation indicates that the higher the uniqueness of the consumption signature, the greater the classification accuracy. In Figure 6, each plot represents a single instance from the top three matched classes. Clearly, each of these signatures exhibits significant dissimilarity from the others. Remarkably, these consumption patterns tend to recur consistently within their respective classes, thereby aiding in efficient identification processes. Figure 7 showcases the classes "Fan," "Hairdryer," and "Heater," respectively.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 9 Issue 37 (December 2024) PP. 133-146
DOI: 10.35631/JISTM.937011

It is evident that, despite belonging to different classes, the consumption signatures exhibit remarkable similarity, primarily differing in amplitudes. This poses a substantial challenge in the identification process, leading to false positives and consequently reducing the accuracy of the final similarity score. This argument is reinforced through a detailed examination of the classification results. For instance, considering sample 113 from the "Fan" class (Figure 7(a)), the final similarity score indicates that "Air Conditioner" has the best match at 24.34, while "Fan" is the second-best match at 24.43. The marginal difference in similarity scores, only 0.09, can be ascribed to the occurrence of these false positives. It's worth noting that other classes, like "Air Conditioner," also generate comparable signatures of the power consumptions, contributing to the challenge of accurate classification.

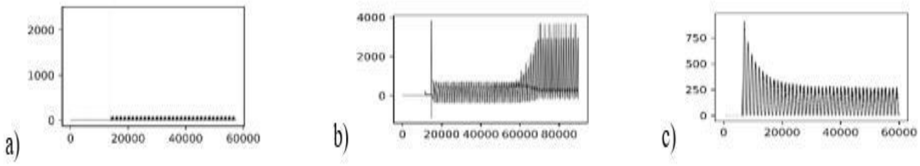

Figure 6: Power Consumption Plots for (a) “CFL”, (b) “Microwave”, (c) “Vacuum”

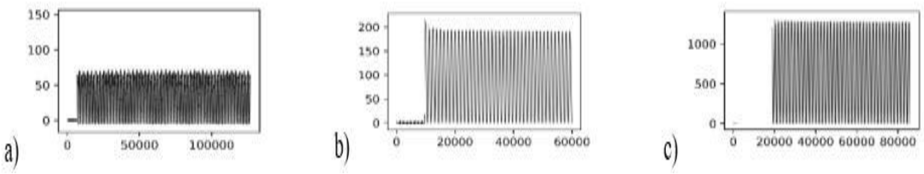

Figure 7: Power Consumption Plots for (a) “Fan”, (b) “Hairdryer, (c) “Heater”

Essentially, the dynamic data reduction approach has demonstrated notable improvements in appliance identification for Non-Intrusive Load Monitoring (NILM). It effectively and accurately extracts relevant regions of interest from lengthy time series. Furthermore, it provides the value for window length, a critical parameter required for Matrix Profile (MP) computation. These identified regions significantly enhance appliance identification compared to full time series comparisons. These observations are further reinforced by the consistency of results across experiments conducted under the same settings. The baseline experiment yielded an accuracy of 25.45%. However, with the implementation of the dimension reduction technique, the accuracy increased substantially to 69.09%, marking a significant improvement.

Conclusion

The proposed dynamic data reduction methodology and the use of the Matrix Profile technique successfully improved the appliance classification accuracy in Non-Intrusive Load Monitoring (NILM) applications. Experimental results demonstrated a substantial increase in accuracy when employing the proposed approach. However, the study faced some limitations, including challenges with overlapping and similar appliance signatures, which occasionally resulted in misclassification. Additionally, the reliance on specific datasets and the exclusion of appliance classes may have constrained the generalizability of the findings. Future works could explore clustering techniques in gathering the insights of the dominant time series pattern within each class. This has the potential to pave the way for further advancements in the appliance identification domain.

Acknowledgements

This work is supported by: Ministry of Higher Education (MoHE), Fundamental Research Grant Scheme (FRGS), Grant No: FRGS/1/2020/ICT02/SWIN/03/3

References

Á. Hernández, et al. (2019). Applications of NILM Techniques to Energy Management and Assist-ed Living. IFAC-PapersOnLine. 52, 11, 164-171.

Angelis, et al. (2022). "NILM applications: Literature review of learning approaches, recent developments and challenges." Energy and Buildings 261: 111951.

Basu, et al. (2016). "A comparative study of low sampling non intrusive load dis-aggregation." In IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society, pp. 5137-5142. IEEE.

Bouhouras, et al. (2019). "A NILM algorithm with enhanced disaggregation scheme under harmon-ic current vectors." Energy and Buildings 183: 392-407.

Dash, Suryalok, and N. C. Sahoo. (2022). "Electric energy disaggregation via non-intrusive load monitoring: A state-of-the-art systematic review." Electric Power Systems Research 213: 108673.

Deppe, et al. (2022). "State of the Art in Time Series Motif Discovery." Discovery of Ill–Known Motifs in Time Series Data: 29-43.

Hafiz K. Iqbal, et al., (2021). A critical review of state-of-the-art non-intrusive load monitoring datasets. Electric Power Systems Research. 192, 106921

Himeur, Yassine, et al. (2022). "Recent trends of smart nonintrusive load monitoring in buildings: A review, open challenges, and future directions." International Journal of Intelligent Systems 37, no. 10: 7124-7179.

Jaradat, Abdelkareem, et al. (2020). "Demand response for residential uses: A data analytics approach." In 2020 IEEE 6th World Forum on Internet of Things, pp. 1-6. IEEE.

Kang, Hyoeun, and Howon Kim. (2020). "Household appliance classification using lower odd-numbered harmonics and the bagging decision tree." IEEE Access 8: 55937-55952.

Linardi, Michele, et al. (2018). "Matrix profile X: VALMOD-scalable discovery of variable-length motifs in data series." In Proceedings of the 2018 International Conference on Management of Data, pp. 1053-1066.

Malik, H. M., Salem, N., & AlSabban, M. S. (2021, December). A Comparative Study on the Performance of Hidden Markov Model in Appliance Modeling. In *2021 5th International Conference on Power and Energy Engineering (ICPEE)* (pp. 145-152). IEEE.

Medico, Roberto, et al. (2020). "A voltage and current measurement dataset for plug load appliance identification in households." Scientific data 7, no. 1: 49.

Miquey, N., & Grover-Silva, E. (2021). Non-Intrusive Load Monitoring of Single and Aggregated Profiles with a Hidden Markov Model. In *ENERGY 2021*.

Moradzadeh, Arash, et al. (2020). "Improving residential load disaggregation for sustainable development of energy via principal component analysis." Sustainability 12, no. 8: 3158

Rana, B., Singh, Y., & Singh, P. K. (2021). A systematic survey on internet of things: Energy efficiency and interoperability perspective. *Transactions on Emerging Telecommunications Technologies*, *32*(8), e4166.

Sankara, A. (2014). *Energy disaggregation in NIALM using hidden Markov models*. Missouri University of Science and Technology

Sarhan, Amany M. (2023). "Data Mining in Internet of Things Systems: A Literature Review." Journal of Engineering Research 6, no. 5: 252-263.

Tezde, Efe Isa, and Eray Yildiz. (2022). "A Comprehensive Survey for Non-Intrusive Load

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 9 Issue 37 (December 2024) PP. 133-146
DOI: 10.35631/JISTM.937011
Monitoring." Turkish Journal of Electrical Engineering and Computer Sciences 30, no. 4: 1162-1186.

Wójcik, Augustyn, et al. (2021). "Identification of the State of Electrical Appliances with the Use of a Pulse Signal Generator." Energies 14, no. 3: 673.

Wu, Z., Wang, C., Zhang, H., Peng, W., & Liu, W. (2021). A time-efficient factorial hidden Semi-Markov model for non-intrusive load monitoring. *Electric Power Systems Research*, *199*, 107372.

Wu, Z., Wang, C., Peng, W., Liu, W., & Zhang, H. (2021). Non-intrusive load monitoring using factorial hidden markov model based on adaptive density peak clustering. *Energy and Buildings*, *244*, 111025.

Yang, Chen. (2021). "A novel uncertainty-oriented regularization method for load identification." Mechanical Systems and Signal Processing 158: 107774

Zhu, et al. (2016). "Matrix profile I: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets." In 2016 IEEE 16th international conference on data mining (ICDM), pp. 1317-1322. IEEE.

Zhu et al. (2016). "Matrix Profile II: Exploiting a Novel Algorithm and GPUs to Break the One Hundred Million Barrier for Time Series Motifs and Joins", 2016 IEEE 16th International Conference on Data Mining (ICDM).

Zymbler, at el. (2021). "Matrix profile-based approach to industrial sensor data analysis inside RDBMS." Mathematics 9, no. 17: 2146.