

Profiling Household Appliance Electricity Usage with N-Gram Language Modeling

Daoyuan Li, Tegawendé F. Bissyandé, Sylvain Kubler, Jacques Klein and Yves Le Traon
Interdisciplinary Centre for Security, Reliability and Trust
University of Luxembourg
4, rue Alphonse Weicker, L-2721 Luxembourg
Email: {firstname.lastname}@uni.lu

Abstract—Household appliance classification, based on electricity usage patterns, is gaining a momentum in an era where energy saving has become a priority and connected objects are leveraged to influence consumers’ behaviors. In this respect, electricity usage profiling of household appliances is an important step for identifying malfunctioning devices and generating automatic alerts about unusual consumptions. To support the implementation of such profiling tasks in a practical and scalable way, a language modeling approach for classifying time series representing appliances’ energy consumption readings is developed in this paper. This approach captures both local discriminative features of consumption profiles and overall curve shapes to yield high classification accuracy, even in the presence of unavoidable noise. Our approach is evaluated against state-of-the-art approaches, and validated based on a large public dataset from 27 different households with energy readings during a one year span. Our experiment findings also show that how the approach remains efficient even in the presence of noise.

I. INTRODUCTION

Citizens in modern societies are increasingly aware of the urgent need to make better use of energy resources and move to more sustainable and greener development paths. There are some signs that the rate of improvement in energy efficiency has been increasing slightly in the last few years, but much remains to be done, especially to turn consumers’ awareness into concrete and tangible actions. Typically, while one of the major sources of electric consumption is household appliances, studies show that inhabitant behaviors have not yet fully shifted [1]. The introduction of smart meters in households is predicted as a defining factor to change such behaviors and incur reduction of energy usage [2], as they are promising means for gaining insights on energy consumption by mining electricity usage patterns.

Smart meters measure energy consumption and provide online readings on electricity usage from households, thus allowing a fine-grained visualization of real-time consumption. This data can be leveraged to detect malfunctioning appliances or discover abnormal situations once usage patterns of appliances have been identified. In order to avoid reluctance and involve the general public in the usage pattern identification task, the research community is investigating automated processes (e.g., to prevent inhabitants being requested to manually indicate which appliance is generating a given consumption signal). In this regard, household appliance classification has become increasingly popular among the research community. The development of Non-Intrusive Load Monitoring (NILM) techniques [3] has made it possible, in a non-intrusive manner,

to disaggregate households’ single-point energy consumption measurements into individual devices’ consumption. Analyzing the usage patterns of different household devices helps, among other things, to better predict future consumptions and ease the macro-management of overall power demand for smart grids [4].

State-of-the-art research works in this area often apply time series classification techniques for identifying appliance electricity usage profiles [5], [6]. However, our research claims that such approaches are not suitable for large-scale electricity usage due to their time complexity, and the fact that they may appear sensitive to noises (leading to inaccuracies in classification of similar appliances). One research hypothesis of this study is that electricity usage of a given appliance can be viewed as unique and recognizable repeating signals that could be represented as a sequence of *sentences*. In this respect, our research aims to develop an innovative time series classification approach based on language modeling techniques that help to identify the “language” of each type of appliances. In a more concrete level, the language modeling approach aims to capture with its n -grams both local discriminative features and overall curve shapes. On the one hand, local discriminative features (e.g., sharp edges) help to distinguish/identify appliances with different electricity usage patterns during turn-on and turn-off phases. On the other hand, overall curve shapes are essential to minimize the impact of noise, which can affect local features. Existing techniques [7] often focus on detecting whether a given appliance is on or off, and at what time it has been switched. As a preliminary investigation in the direction of energy disaggregation, our study assesses the potential of the proposed language modeling approach to identify the contributions of several appliances in a combined electricity usage data, and particularly to identify whether pairwise combinations of appliance consumption data can be accurately profiled. Overall, the paper’s contributions are:

- Proposal and development of an innovative accurate and practical approach to profiling household appliances based on their electricity consumption readings.
- Evaluation – *using a large dataset* – of our approach against the state-of-the-art approach to time series classification.
- Presentation of preliminary findings on profiling of electricity consumption of combinations from different household appliances.

Related work in the literature and the necessary background for the approach are respectively introduced in Sections II

and III. Section IV provides greater detail on our language modeling approach, while experiments and evaluation results are presented in Section V; discussion and conclusion follow in Sections VI and VII.

II. RELATED WORK

A set of techniques for appliance classification and profiling have been introduced in the literature. For example, NILM systems break summarized appliance usage information into individual appliance usage patterns. Typically, household appliances can be divided into four categories: 1) permanent devices that work constantly with steady active and reactive power consumption; 2) on-and-off devices that could be either on or off at any point of time (e.g., lights, toasters...); 3) appliances with many states where energy consumption may switch from one state to another (those appliances can be modeled as Finite State Machines – FSM); and 4) continuously variable devices, which exhibit no usage patterns. Given these categories, only categories 2) and 3) can be detected using NILM techniques [7]. Indeed, the main NILM research effort has been on appliance signature extraction, which is often viewed as a pattern recognition task. Numerous techniques have been applied for that purpose, including k Nearest Neighbors (kNN) [8] that looks for the most similar patterns from the existing pattern pool, Naïve Bayes classifier [5] that is a probabilistic classifier modeling patterns as feature vectors (assuming feature independence within vectors), and Powerlets [6] that uses *sharp edges* in appliance usage patterns as discriminative features. It is important to note that some recent NILM research takes advantage of features other than appliances’ real and reactive power (e.g., transients and harmonics) [9], however, such techniques require very high sampling frequency leading to expensive measurement hardware.

Regarding appliance usage representation, many research efforts have adopted the time series model, which is the “natural” representation of device energy consumption data. Lines et al. [5] classify household devices by electricity usage profiles using a time series classification approach, and compare classification results using different algorithms such as kNN, Naïve Bayes, Random Forest, *etc.* Similarly, Basu et al. [10] adapted a time series model by introducing a multi-label classifier approach to predict appliance usage in the near future.

Time series classification is an active research field in the machine learning community, where a number of approaches and algorithms have been proposed, including neural networks [11], decision trees [12] and SVM [13], while empirical studies have shown that kNN works exceptionally well [14]. To perform best, k NN classifiers leverage the Dynamic Time Warping (DTW) distance, thus mitigating problems by warping the time axis [15]. Another line of research – *shapelets-based classifiers* [16] – focuses on finding the most discriminative features. *Shapelet* algorithms are proven to be accurate for time series classification tasks, but they are generally time consuming due to the exhaustive process searching for the best features. As a result, latest research efforts focus on reducing the time complexity of these algorithms [17], [18].

III. BACKGROUND

In this study, household appliance electricity consumption is also modeled as time series, which consist of ordered

data sequences. The grey curve in Fig. 1 illustrates a time series sample representing the energy consumption pattern of a fridge (over a period of time of 200 minutes). From a human being perspective, the corresponding pattern can be identified: *30-minute low consumption, followed by 10-minute medium consumption, then another 30-minute low consumption, etc.* However, from a computer perspective, the real-valued data may cause confusions and noises since minor variations may not be relevant to the overall picture. As a result, we use Symbolic Aggregate approXimation (SAX) [19] to both remove noises and reduce the length of the series, which strongly reduces the processing complexity. SAX reduces the dimension of the time axis to a fraction of the original series (dimension reduction), and the dimension of the vertical axis from infinite to a discrete and finite number depending on the user-specified alphabet size (numerosity reduction). The black dashed curve in Fig. 1 illustrates the approximated representation of the original data, and the alphabets represents the numerosity reduction results: the 200-minute fridge consumption data is reduced into a string of 20 alphabet letters: *aaacbabdddcaaceebdd*.

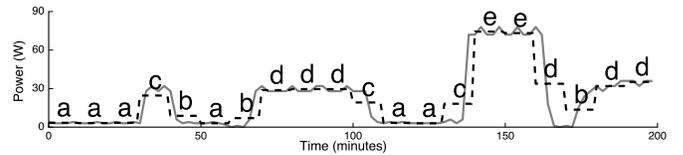


Fig. 1: Original electricity consumption data of from a fridge (grey curve) and corresponding SAX representation using a five-letter alphabet. Note that the gap between *d* and *e* is much bigger than that the one between *a* and *b*, because SAX works on value distributions instead of arbitrary ranges.

After obtaining the string representations of time series, strings of a same class – i.e., string representations of electricity consumption data from the same household appliance – can be incorporated into one set, based on which we generate n -gram language models. A language model can be defined as a statistical distribution of all words in a corpus [20]. Given a sequence $S = s_1, s_2, \dots, s_n$, it has a probability score of $P(S) = P(s_1, s_2, \dots, s_n)$, which is calculated as in Equation 1. A unigram model approximates this score by assuming complete independence among all subsequences, which can be formalized as in Equation 2.

$$P(s_1, s_2, \dots, s_n) = P(s_1)P(s_2|s_1)\dots P(s_n|s_1, \dots, s_{n-1}) \quad (1)$$

$$P(S) = P(s_1, s_2, \dots, s_n) \approx \prod_{i=1}^n P(s_i) \quad (2)$$

Bigram models take into account the likelihood of one subsequence followed by another (i.e., the conditional probabilities of one subsequence given the previous one). Bigram models are more accurate than unigram models at the expense of storing more probability entries. In English *eat* and *drink* are unigram examples that may have similar probabilities, but *eat bread* should have a much higher probability than *drink bread*. The probability score of a sequence following a bigram model is computed by Equation 3.

$$P(S) = P(s_1, s_2, \dots, s_n) \approx \prod_{i=1}^n P(s_i|s_{i-1}) \quad (3)$$

IV. INNOVATIVE APPROACH FOR HOUSEHOLD APPLIANCE PROFILING

As previously stated, our approach is based on the research hypothesis that different household appliances exhibit different electricity consumption profiles in similar ways as how people speak different languages. That is, each natural language has its own corpus and grammar information, which roughly corresponds to the unigram and bigram language models acquired from the symbolized electricity consumption readings. Unlike powerlets [6] that builds a *dictionary* containing discriminative consumption signatures for different devices (i.e., definitive sharp edges in electricity consumption curves), our approach considers both local discriminative features and overall curve shapes. Thus our approach differs from powerlets in the following aspects:

- Powerlets is, in essence, an application of shapelet algorithms to appliance classification tasks, thus suffering from the same shapelets algorithm issues (notably time-consuming due to exhaustive search for discriminative features). As a result, powerlets may not be suitable for large-scale industrial applications.
- Powerlets can be too fine-grained and not resistant to noises due to the lack of global contexts. For example, powerlets may not be able to classify/cluster two hair driers of different power draws in a same category since they will have different height in sharp edges.
- Powerlets requires a high data sampling rate for the reason that sharp edges would get the same (or similar) sharpness and/or steepness. This requires more expensive measurement hardware and limits powerlets' application scope.

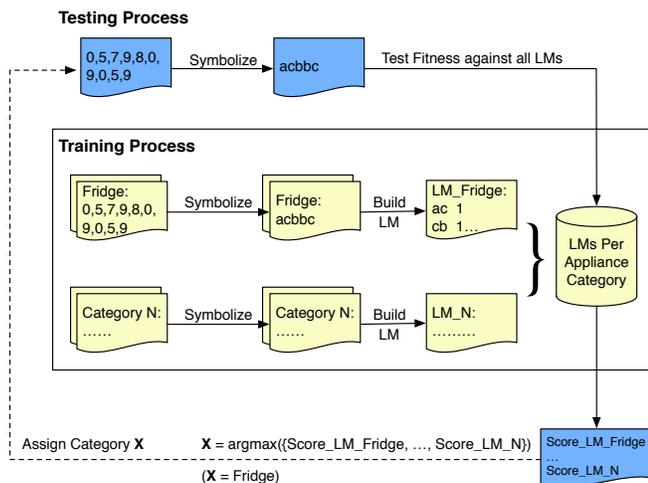


Fig. 2: Illustrative steps for learning appliance profiles and classifying an instance from an unknown category.

Fig. 2 provides insight into the overall process underlying our approach. The final output of the training process is a unigram and bigram model linked to each category, whose respective steps to obtain those models (i.e., profiles) are:

- 1) Transform all real-valued electricity consumption readings into strings in order to reduce data dimension and processing complexity. SAX is used for this purpose.

- 2) Extract subsequences from each consumption string and build unigram and bigram language models for each appliance category. Since these strings have no word boundaries, we extract subsequences based on a sliding window mechanism. For instance, using a sliding window of size 3, the string `acbbc` can be broken down as follows: `acb`, `cbb` and `bbc`. Then we count each subsequence's number of occurrences and convert them into probabilities to build unigram models. Likewise, the probabilities of bigrams and build bigram models are extracted.
- 3) Finally, step 2) is repeated to build language models for each appliance category.

Profiles obtained from the training phase are stored into a database, so that they can be retrieved later when assigning profiles to unlabeled consumption readings. To be precise, the following steps are followed when classifying an energy consumption data from an unknown type of appliance:

- 1) Transform this piece of data into a string following the same settings as in step 1) of the profile acquisition phase.
- 2) Test the string against all available appliance category models (profiles) to produce a fitness score with respect to each model. Such a fitness score is calculated as the maximum probability of dividing this sample using the language model.
- 3) Compare fitness scores with respect to all profile models. Generally, the maximum probability from all models implies the best fitness, thus the appliance represented by this specific model is the most likely appliance that has produced the electricity consumption footprint.

The fitness score of one piece of symbolized data $D_{1 \rightarrow n} = d_1, d_2, \dots, d_n$, with respect to a profile model M , is calculated by recursively finding possible candidates to split the data according to M , as formalized in Equation 4 (with $1 < k < n$ and $D_{1 \rightarrow k} \in M$). This process is repeated for all other segments. The fitness score can therefore be defined as the maximum probability score, as formalized in Equation 5.

$$C(D, M) = D_{1 \rightarrow k} \cup C(D_{k+1 \rightarrow n}, M) \quad (4)$$

$$F(D, M) = \max_{c \in C(D, M)} \prod_{i=1}^{|c|} P(c_i | c_{i-1}) \quad (5)$$

It is worth noting that the recursive process for enumerating possible ways of splitting data samples can become time-consuming. However, our approach leverage dynamic programming to store recurring segments and to reduce the computational complexity to roughly $O(n)$, similar to or better than state-of-the-art approaches.

V. EMPIRICAL EVALUATION

Our approach has been implemented in order to evaluate the classification performance against state-of-the-art approaches. The set of experiments has been conducted using the computing platform presented in [21], and considering several datasets and scenarios. Unless otherwise specified, we extracted – *for all datasets* – subsequences of lengths ranging from 2 to 20. Furthermore, data samples longer than 100 are reduced to 100 using SAX in order to speed up experiments. The evaluation results and findings are presented and discussed through the

TABLE I: Characteristics of appliance electricity usage data from the UCR archive and classification accuracy comparison between the performance of 1NN with Euclidean and DTW distance and our approach. Best classification accuracy results are highlighted in bold font.

Dataset Name	# Classes	Characteristics			Instance Length	Classification Accuracy		
		# Training Instances	# Testing Instances	Instance Length		1NN (Euclidean)	1NN (DTW Best Warping)	Our Approach
Computers	2	250	250	720	0.576	0.62	0.668	
ElectricDevices	7	8,926	7,711	96	0.55	0.624	0.651	
LargeKitchenAppliances	3	375	375	720	0.493	0.795	0.72	
RefrigerationDevices	3	375	375	720	0.395	0.44	0.528	
ScreenType	3	375	375	720	0.36	0.411	0.448	
SmallKitchenAppliances	3	375	375	720	0.341	0.672	0.659	

sections V-A to V-C. The source code of our prototype is released to the public¹ to increase reproducibility.

A. Evaluation against normalized datasets

Our approach is compared against state-of-the-art approaches with an openly accessible dataset archive (the UCR archive [22]), which is widely adopted in the research community. Besides, it comes with classification results for state-of-the-art classification algorithms for benchmark comparison, namely Euclidean- and DTW-distance based Nearest Neighbor as shown in TABLE I. This archive contains many different datasets from interspecies images to medical applications such as electrocardiograms.

In this research, only datasets related to household appliance electricity usage profiles are tested since our main interest is on classifying electric devices. The characteristics of these datasets are summarized in TABLE I, where performance of our approach is compared with Euclidean-based 1NN and DTW-based 1NN (best accuracy varying the warping size from 0 to 100 percent of the series length, while best accuracy varying SAX alphabet size parameter from 3 to 20). Although our approach performs well with regard to these datasets, a more in-depth study should be conducted due to the fact that data that compose these datasets have already been normalized, adding that each dataset contains only a small number of samples as well as classes. Furthermore, the training sets and testing datasets have been roughly equally divided in terms of number of instances.

B. Evaluation against real-world readings

To test our approach with real-world data, the dataset obtained from the UK Household Electricity Usage Survey (HEUS) project [23] has been used, which contains electric appliance usage readings from 251 households (monitoring periods: 2010 to 2011). In our study, a subset containing more than 50 million appliance electricity usage readings from 27 distinct household have been considered (readings being sampled at two-minute intervals). As a first step, from this subset the ten appliances that have the most recorded readings has been extracted, along with their electricity consumption patterns as depicted in Fig. 3.

Afterwards, readings are clustered into time series with size of 100, which corresponds to appliances' electricity readings during a 200-minute period. Note that readings that are roughly constant during each that period have been removed, leading to 25,652 time series. A ten-fold cross-validation is then performed to classify these time series. More specifically, this

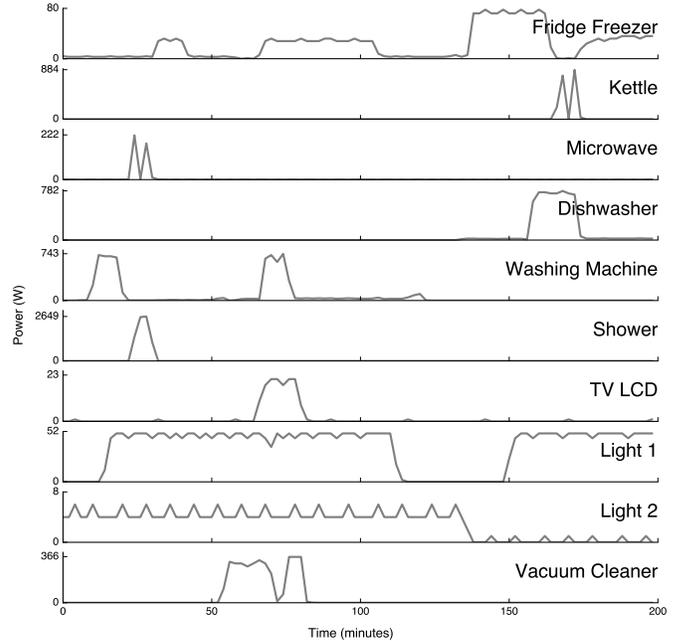


Fig. 3: Electricity consumption patterns of ten most recorded appliances from 27 households surveyed in the HEUS project.

dataset is firstly divided into ten portions, nine of them being used for training and the remaining one being used for testing. This process is repeated ten times and the average classification accuracy is reported at 77.4%. Compared with [5], where data is sampled with a 15-minute interval and clustered into weekly and daily series, the best results obtained are 61.34% and 55.81% respectively. In addition, since we have mixed appliance usage data across households, the good classification results suggests that certain kinds of household appliances exhibit similar profiles. Thus, it is sensible to build a profile repository for different appliances so that, later, unknown appliances can potentially be matched with profiles in the repository, without the need of setting up monitoring devices in each household and learning device profiles from scratch.

TABLE II presents the confusion matrix from the ten-fold cross-validation experiment. The standard F-measure scores (the harmonic mean of precision and recall calculated following Equation 6) are also reported in order to present the performance of our approach in a more interpretable manner.

$$F - measure = F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (6)$$

It can be observed that our approach provides good results for certain appliance categories such as *Fridge Freezer* and

¹<https://github.com/serval-snt-uni-lu/profiling-appliances>

TABLE II: Confusion matrix from ten-fold cross-validation experiment on the most recorded appliances.

		Classified As										F-Measure
		Fridge Freezer	Kettle	Microwave	Dishwasher	Washing Machine	Shower	TV LCD	Light 1	Light 2	Vacuum Cleaner	
Actual	Fridge Freezer	9,380	1	7	97	58	1	509	40	16	1	0.931
	Kettle	3	4,343	215	21	90	109	6	0	6	267	0.831
	Microwave	36	885	936	64	171	242	29	7	6	164	0.482
	Dishwasher	20	23	35	895	203	28	12	9	3	32	0.660
	Washing Machine	46	22	58	159	1,302	48	53	7	3	32	0.676
	Shower	0	36	16	27	35	246	2	5	4	29	0.408
	TV LCD	476	20	24	99	186	37	2,174	116	106	22	0.699
	Light 1	52	5	13	27	28	13	101	341	40	10	0.575
	Light 2	16	8	8	18	15	34	70	32	227	2	0.540
	Vacuum Cleaner	1	52	33	46	32	48	3	0	0	45	0.104

Kettle. Furthermore, our approach tends to perform better for appliances with larger number of reading records, indicating that good performance can be achieved with sufficiently large training dataset.

C. Combining appliance consumption readings

So far, the evaluation focused on exclusive energy readings for specific appliances. Unfortunately, in many cases, it is often difficult to get such readings. Instead, it is likely that some appliances (e.g., a fridge) stay on the ON mode, while others more occasionally used appliance (e.g., a microwave) is turned ON or OFF. Thus, a deeper investigation is carried out in this section in order to classify the combinations of appliances' energy consumptions. This study will help us to gain insights on the potential of the language modeling approach to support energy disaggregation schemes.

To this end, we first find out the ten most recorded appliances within the dataset following the same precedures mentioned previously. Then, we investigate which of these ten appliances are found in each household and sum up the power consumption data based on timestamps. In this way we are able to keep the original appliance usage patterns – i.e., which two appliances are used within the same period – from each household. Afterwards, we find out the most frequent consumption sums representing appliance usage patterns and set to classify them using our language modeling based approach. All the combination samples are mixed in order to create a sufficiently large dataset. Finally, a ten-fold cross-validation experiment is conducted in order to differentiate various appliance usage consumption combinations.

TABLE III presents the most frequently found appliance combinations and classification results. As shown, it is possible to differentiate combinations of appliance usage, e.g., the *Fridge Freezer* and *Light 1* combination can be differentiated from the *Kettle* and *Microwave* combination. This can be explained by the fact that each appliance has a different usage patterns (*Fridge Freezer* and *Light 1* tend to be constant while *Kettle* and *Microwave* both exhibit burst patterns). On the other hand, some combinations such as the *Fridge Freezer* and *Light 1* vs. *Fridge Freezer* and *Washing Machine* are difficult to differentiate, whose possible reasons may be:

- Appliance usage data are sampled at a low frequency (a 2 minute interval may not be enough to pertain appliance usage characteristics, or to sum up the usage data).
- Our approach reduces real-valued power consumption into alphabets, and this numerosity reduction process may have incurred inaccurate data representation for combinations.

TABLE III: Classification results for ten-fold cross-validation between two appliance energy consumption combinations. Columns P, R and F stand for Precision, Recall and F-Measure respectively.

Combination 1	Combination 2	P	R	F
Fridge Freezer + Kettle	Kettle + Light 1	0.941	0.941	0.941
Fridge Freezer + Kettle	Kettle + Microwave	0.959	0.958	0.958
Fridge Freezer + Light 1	Fridge Freezer + Microwave	0.551	0.549	0.549
Fridge Freezer + Light 1	Fridge Freezer + Kettle	0.647	0.637	0.638
Fridge Freezer + Light 1	Kettle + Light 1	0.977	0.977	0.977
Fridge Freezer + Light 1	Kettle + Microwave	0.990	0.990	0.990
Fridge Freezer + Light 1	Fridge Freezer + Washing Machine	0.472	0.469	0.470
Fridge Freezer + Microwave	Fridge Freezer + Kettle	0.620	0.617	0.617
Fridge Freezer + Microwave	Kettle + Light 1	0.975	0.975	0.975
Fridge Freezer + Microwave	Kettle + Microwave	0.987	0.987	0.987
Fridge Freezer + Washing Machine	Fridge Freezer + Microwave	0.538	0.538	0.538
Fridge Freezer + Washing Machine	Kettle + Microwave	0.990	0.990	0.990
Fridge Freezer + Washing Machine	Kettle + Light 1	0.977	0.977	0.977
Fridge Freezer + Washing Machine	Fridge Freezer + Kettle	0.600	0.599	0.598
Kettle + Light 1	Kettle + Microwave	0.618	0.596	0.587
Kettle + Washing Machine	Kettle + Microwave	0.625	0.573	0.558
Kettle + Washing Machine	Kettle + Light 1	0.585	0.576	0.574
Kettle + Washing Machine	Fridge Freezer + Microwave	0.979	0.979	0.979
Kettle + Washing Machine	Fridge Freezer + Light 1	0.986	0.986	0.986
Kettle + Washing Machine	Fridge Freezer + Washing Machine	0.961	0.961	0.961
Kettle + Washing Machine	Fridge Freezer + Kettle	0.949	0.948	0.948
Microwave + Light 1	Fridge Freezer + Washing Machine	0.965	0.965	0.965
Microwave + Light 1	Fridge Freezer + Light 1	0.961	0.961	0.961
Microwave + Light 1	Fridge Freezer + Kettle	0.947	0.947	0.947
Microwave + Light 1	Fridge Freezer + Microwave	0.955	0.955	0.955
Microwave + Light 1	Kettle + Microwave	0.839	0.833	0.829
Microwave + Light 1	Kettle + Washing Machine	0.863	0.850	0.847
Microwave + Light 1	Kettle + Light 1	0.798	0.795	0.792

- The size of training data may not be sufficiently large, thus leading to insufficient training process and misclassified testing instances.

VI. DISCUSSIONS

The approach proposed in this paper relies on a mechanism to transform household appliance electricity usage data into a string, where strings of a same appliance category forms a appliance-specific dictionary. The results from our empirical experiments showed how performant our approach is, and made it possible to validate the analogy between appliance energy consumption patterns and natural language sentences. Besides classification accuracy, this approach has a few other advantages. For example, our approach is able to handle data samples of different length. This is helpful in that it eases data preparation and segmentation in practice. Furthermore, this approach makes it easy to store data samples as well as learned profiles/models in a distributed manner, so that existing distributed and cloud computing techniques can be easily taken advantage of and further boost the efficiency of our approach.

Nonetheless, some limitations of our approach can be discussed. First, the model-based nature underlying our approach demands a large training set so as to achieve a high performance. Then, with our current settings, $T_1 = [1, 1, 4, 4, 10, 10]$ and $T_2 = [100, 100, 300, 300, 600, 600]$ will be transformed into the same string (*aabbcc*) when SAX alphabet size is set to 3

and string length set to keep the original series length, which can be an issue and cause mis-profiling of appliances. On the opposite, this issue can benefit our approach when profiling a same category of appliances having different power draws. For example, a 1000 KW and 1500 KW hair dryer should indeed have same consumption patterns, while a 1000 KW dishwasher should exhibit different profiles compared with the 1000 KW hair dryer. For the same reason, when combining appliance consumption readings, data from one appliance with small power draw may be dwarfed by an appliance with a much larger power draw, so that the former becomes noises and the overall curve shape is dominated by the latter.

VII. CONCLUSIONS AND FUTURE WORK

Electricity usage profiling of household appliances is becoming an important step for identifying malfunctioning devices and generating automatic alerts about unusual consumptions. In this context (i.e., household appliances profiling context), this paper investigates the capability of a language modeling approach for time series classification. To this end, an innovative approach is proposed in this paper, which aims to first transform energy consumption readings – *which consist of real-valued time series data* – into texts, and then to build per-class language models (i.e. *profiles*) from these texts. Such class models can therefore be used for new electricity usage readings in order to predict the corresponding appliance category. The proposed approach has been implemented and evaluated through a set of experiments considering both normalized datasets from the research community and real world datasets from the UK Household Electricity Usage Survey project (27 households monitored over one year). These experiments show that our approach performs generally better than state-of-the-art time series classification approaches.

ACKNOWLEDGMENT

The authors would like to thank the Ministry of Economy Luxembourg, Paul Wurth S.A. and the National Research Fund Luxembourg (grant 9095399) for sponsoring this research work.

REFERENCES

- [1] J. Morales Pedraza, *Electrical Energy Generation in Europe: The Current Situation and Perspectives in the Use of Renewable Energy Sources and Nuclear Power for Regional Electricity Generation*, Springer, Ed. Springer, 2015.
- [2] S. Darby, “Making it obvious: designing feedback into energy consumption,” in *Energy efficiency in household appliances and lighting*. Springer, 2001, pp. 685–696.
- [3] G. W. Hart, “Nonintrusive appliance load monitoring,” *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.
- [4] A. Barbato, A. Capone, M. Rodolfi, and D. Tagliaferri, “Forecasting the usage of household appliances through power meter sensors for demand management in the smart grid,” in *IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, 2011, pp. 404–409.
- [5] J. Lines, A. Bagnall, P. Caiger-Smith, and S. Anderson, “Classification of household devices by electricity usage profiles,” in *Intelligent Data Engineering and Automated Learning-IDEAL 2011*. Springer, 2011, pp. 403–412.
- [6] E. Elhamifar and S. Sastry, “Energy disaggregation via learning powerlets and sparse coding,” in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
- [7] M. Zeifman and K. Roth, “Nonintrusive appliance load monitoring: Review and outlook,” *IEEE Transactions on Consumer Electronics*, pp. 76–84, 2011.
- [8] M. E. Berges, E. Goldman, H. S. Matthews, and L. Soibelman, “Enhancing electricity audits in residential buildings with nonintrusive load monitoring,” *Journal of industrial ecology*, vol. 14, no. 5, pp. 844–858, 2010.
- [9] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong, “Power signature analysis,” *Power and Energy Magazine, IEEE*, vol. 1, no. 2, pp. 56–63, 2003.
- [10] K. Basu, V. Debusschere, and S. Bacha, “Appliance usage prediction using a time series based classification approach,” in *Thirty-Eighth Annual Conference of the IEEE Industrial Electronics Society (IECON)*. IEEE, 2012, pp. 1217–1222.
- [11] A. Nanopoulos, R. Alcock, and Y. Manolopoulos, “Feature-based classification of time-series data,” *International Journal of Computer Research*, vol. 10, no. 3, 2001.
- [12] J. J. Rodríguez and C. J. Alonso, “Interval and dynamic time warping-based decision trees,” in *Proceedings of the 2004 ACM symposium on Applied computing*. ACM, 2004, pp. 548–552.
- [13] Y. Wu and E. Y. Chang, “Distance-function design and fusion for sequence data,” in *Proceedings of the thirteenth ACM international conference on Information and knowledge management*. ACM, 2004, pp. 324–333.
- [14] G. E. Batista, X. Wang, and E. J. Keogh, “A complexity-invariant distance measure for time series,” in *Proceedings of the Eleventh SIAM International Conference on Data Mining (SDM)*, vol. 11, 2011, pp. 699–710.
- [15] C. A. Ratanamahatana and E. Keogh, “Three myths about dynamic time warping data mining,” in *Proceedings of the 2005 SIAM International Conference on Data Mining (SDM05)*, 2005, pp. 506–510.
- [16] L. Ye and E. Keogh, “Time series shapelets: a new primitive for data mining,” in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 947–956.
- [17] A. Mueen, E. Keogh, and N. Young, “Logical-shapelets: an expressive primitive for time series classification,” in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011, pp. 1154–1162.
- [18] T. Rakthanmanon and E. Keogh, “Fast shapelets: A scalable algorithm for discovering time series shapelets,” in *Proceedings of the thirteenth SIAM conference on data mining (SDM)*, 2013.
- [19] J. Lin, E. Keogh, L. Wei, and S. Lonardi, “Experiencing sax: a novel symbolic representation of time series,” *Data Mining and knowledge discovery*, vol. 15, no. 2, pp. 107–144, 2007.
- [20] J. M. Ponte and W. B. Croft, “A language modeling approach to information retrieval,” in *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, 1998, pp. 275–281.
- [21] S. Varrette, P. Bouvry, H. Cartiaux, and F. Georgatos, “Management of an academic hpc cluster: The ul experience,” in *Proc. of the 2014 Intl. Conf. on High Performance Computing & Simulation (HPCS 2014)*. Bologna, Italy: IEEE, July 2014, pp. 959–967.
- [22] Y. Chen, E. Keogh, B. Hu, N. Begum, A. Bagnall, A. Mueen, and G. Batista, “The ucr time series classification archive,” July 2015, www.cs.ucr.edu/~eamonn/time_series_data/.
- [23] J.-P. Zimmermann, M. Evans, J. Griggs, N. King, L. Harding, P. Roberts, and C. Evans, “Household electricity survey: A study of domestic electrical product usage,” *Intertek Testing & Certification Ltd*, 2012.