Received February 18, 2019, accepted March 21, 2019, date of publication April 11, 2019, date of current version April 22, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2910310

A Modified IP-Based NILM Approach Using Appliance Characteristics Extracted by 2-SAX

YAOXIAN LIU, YI SUN, AND BIN LI
School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China

Corresponding author: Bin Li (18500777272@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51777068.

ABSTRACT Deregulation on the delivery side of the power market has continuously been moving forward worldwide, which make bidirectional flow and interactions between customers and grids needs to be more refined and in-depth. Large-scale coverage of the advanced metering infrastructure (AMI) brings in skyrocketing of an immense amount of fine-grained, real-time consumption data and causes communication traffic congestion between meters and a cloud computing center. To tackle these two challenges, this paper proposes a modified IP-based non-intrusive load monitoring approach using appliance characteristics extracted by quadratic symbolic aggregate approximation (2-SAX). A 2-SAX algorithm is implemented to carry out dimensionality reduction on equipment load data and extracted the state's transition behavior characteristics and operation probability characteristics of each device. The extracted features can use to modify the disaggregation results of integer programming for overcoming the shortcomings of the previous IP approach. The developed method is tested with AMPds dataset. The results of experiments illustrate the 2-SAX consequences in 38.82%, 52.46%, and 13.41% reduction in MAE, MAPE, and RMSE on the heat pump and achieves similar performance on the other appliances, compared with normal SAX. Meanwhile, the proposed method MIP-AC2S delivers significant accuracy advantage and competitive performance over IP, ALIP, and MIP disaggregation method.

INDEX TERMS Non-intrusive load monitoring, appliance characteristics, quadratic symbolic aggregate approximation (2-SAX), load data mining.

I. INTRODUCTION

Two-way flow of load and information, as the Energy Department of US suggested, is one of the core connotations in the construction of smart grid and beneficial to energy utilities to monitor power plants and user status [1]. Advanced Metering Infrastructure (AMI) has gained increasing popularity with the installations in China reached 427 million and non-metering functions of AMI will be vigorously promoted to create advantageous conditions for this bidirectional flow and interactions between customers and grids [2]. Non-intrusive Load Monitoring (NILM), a powerful and cost-effective technique, decomposes the total current or energy consumption measured at the household level into the energy consumption of individual appliances. It helps the Energy Management System (EMS) to gather the clients' load profiles or load patterns [3]. Instead of the aggregated energy consumption, single-appliance consumption level is more conducive to understand users' consumption demand and behaviors [4], which may have great value on accurate demand response (DR) [5], power reliability and efficiency improvement [6], electricity price design [7] and other personalized services [8], [9].

The associate editor coordinating the review of this manuscript and approving it for publication was Ruqiang Yan.

Researches on NILM is a hot spot in load profiling recently. Although high-frequency sampling data leads to higher identification accuracy and more transient features, e.g. harmonics, image information and so on [10], [11], there are a number of problems to be resolved. High-frequency data extraction needs additional measuring equipment on the access line and causes the huge transformation and construction costs [12], [13], which is tricky to be executed. Heretofore, current researches, using hidden Markov model [14],

[15], deep learning [16], [17], non-negative matrix decomposition [3], [18], [19] and other machine learning methods, focus on realizing higher load disaggregation accuracy with low-frequency load data gathered by smart meters. Self-encoding which transforms the load disaggregation problem into a decoding problem using hidden Markov model [14] or deep learning [16], has been verified its effectiveness in sequence-to-sequence method. And Literature [17] introduces deep learning into a sequence-to-point method to obtain better performance. However, machine learning algorithms require a large amount of prior knowledge for training, which has certain limitations. Alternative approaches such as combinatorial optimization or integer programming (IP) have been explored in recent work. Literature [20] proposed an IP-based disaggregation method using high-frequency sampled data. Some enhancements, including additional constraints, correction based on a state diagram, median filtering and linear-programming-based refinement, are implemented in [21] to improve disaggregation accuracy and make it suitable for low-frequency data as well.

There has got some progress in this area recently, however, at least two challenges will be tackled. One challenge is "the deluge of data". The widespread popularity of AMI brings in skyrocketing of an immense amount of fine-grained, real-time consumption data and causes communication traffic congestion between meters and cloud computing center. How to use the redundancy of load data to make a trade-off between acquisition of crucial information and alleviation of communication and storage burden is a prerequisite for the effective analysis of energy consumption data [22], [23]. In this context, data compression algorithms can be well applied to reduce the size of the load data before disaggregation. A variety of mature compression methods such as discrete Fourier transform (DFT) [24], discrete wavelet transform (DWT) [25], Principal Component Analysis (PCA) [26] and Singular Value Decomposition (SVD) [27] have been discussed in the literature. With regard to smart meter data compression, a feature extraction based on non-negative K-SVD algorithm was analyzed in [28] and Symbolic Aggregate Approximation (SAX) is introduced in [29] to reduce the scale of numerical consumption data and formulate the electricity consumption behavior dynamics in adjacent periods. The other challenge is that massive load data should no longer be limited to simple and direct applications but should carry out an in-depth analysis to further utilize hidden value or useful information of that.

To tackle these two challenges, this paper implements a non-intrusive load disaggregation approach, MIP-AC2S, based on modified integer programming using appliance characteristics extracted by quadratic Symbolic Aggregate Approximation (2-SAX).

From the preceding analysis, the contributions of this study are as follows.

1) A 2-SAX is applied instead of SAX to effectively reduce the dimension when the consumption data in different periods differs in magnitude, where first SAX is mainly focuses on determined the magnitude of amplitudes and second one is performed to distinguish load states in same magnitude.

2) MIP-AC2S disaggregation approach is proposed to modify the optimal solution of integrate programming problem using states transition behavior characteristics and operation probability characteristics of each device for overcoming many of the shortcomings of the previous IP-based approach.

3) A Cloud, Edge and End-user synchronizing computing framework combining 2-SAX and MIP-AC2S is proposed to alleviate the transmission pressure on the data link and cloud computing center.

The remainder of this paper is organized as follows. Section 2 briefly introduces the modified IP-based NILM approach using features extracted by 2-SAX algorithm. Section 3 presents the "data deluge" problem and designs a Cloud, Edge and End-user synchronizing computing framework based on edge computing. Section 4 describes the implementation of the numerical experiment and case study using actual AMI data and compares the performance of the proposed MIP-AC2S with previous state-of-the-art methods, namely, the traditional IP algorithm, Aided Linear Integer Programming (ALIP) algorithm proposed in [21] and Mixed-Interger linear Programming (MIP) algorithm proposed by Wittmann F.M. Finally, conclusions are drawn in Section IV.

II. NILM BASED ON MODIFIED IP USING FEATURES EXTRACTED BY 2-SAX

A. BASIC SAX ALGORITHM

SAX is an effective indirect clustering technique for the dimensional reduction of and representation of time series data [30]. Load data from household appliances, typically time series data, are redundant, or rather, most electrical devices are in a stable state over a certain period, during which its consumption will remain constant or fluctuate very little. Consider its lower computational complexity, we select SAX instead of k-means, Mean Shift or Affinity Propagation clustering algorithm to transform the load curve into a discrete symbolic string to reduce the dimension and alleviate the burden between smart meters and data centers.

The following two steps are implemented to discretizes numeric time series into a discrete string:

1. Piecewise Aggregate Approximation (PAA) representation

Consider the actual consumption data $x_j = [x_1, x_2, \cdots, x_M]$, which can be partitioned into H intervals by time domain breakpoint k_i for $i = 0, 1, 2, \cdots, H, H < M$. The PAA representation of x_j is replacing the amplitude fluctuation of ith segment with its average values [31], which can be described as Equation (1):

$$\bar{x}_i' = \frac{1}{k_i - k_{i-1}} \sum_{j=k_{i-1}+1}^{k_i} x_j. \tag{1}$$

It has been proven that averaging of the PAA can weaken the impact of short-duration "spikes" of load profiles.

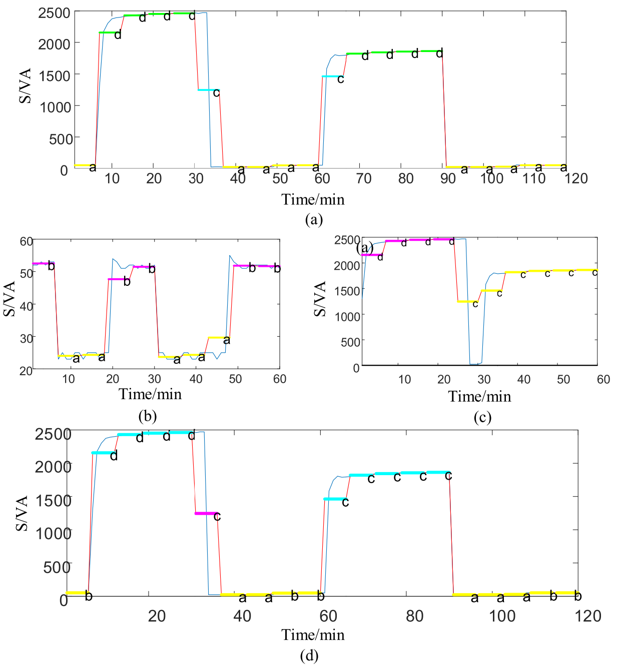

FIGURE 1. **Electricity consumption data of heat pump and its symbolic representation results. (a) using traditional SAX; (b) using 2-SAX for symbol "a" in (a); (c) using 2-SAX for symbol "d" in (a); (d) using 2-SAX for heat pump.**

2. Symbolization using SAX

For symbolizing the PAA representation into a discrete string, the amplitude axis is segment into P intervals and each amplitude range $[\beta_{p-1}, \beta_p]$ corresponds to a symbol ω_p. Therefore, the load curves can be represented by a symbolic string α, as shown in Equation (2):

$$\alpha_i = \omega_p \text{ } if \text{ } \beta_{p-1} < \bar{x}_i^{\prime} < \beta_p. \tag{2}$$

B. 2-SAX FOR HOUSEHOLD APPLIANCE LOAD

Figure 1 (a) shows the electricity consumption data of the heat pump over two hours from AMPds dataset and its symbolic representation with traditional SAX. The time axis is divided into twenty regular periods and these data can be represented as "addddcaaaacddddaaaaa" with three symbols. One of the main concerns of SAX is the determination of the amplitude breakpoint β_p. Traditional SAX, when the load in different periods differs in magnitude, mainly focuses on clusters of different orders of magnitude and cannot distinguish several states of the same order effectively. For example, Symbol "a" in Figure 1 (a) actually represents two load states P1=23VA and P2=52VA, while symbol "d" represents P3=1800VA and P4=2500VA. This may cause a large disaggregation error in NILM and result in unexpected disaggregation deviation for other appliances in the integer programming process.

In order to surmount the limitations of SAX, this paper proposed a secondary clustering algorithm (2-SAX) method. The breakpoint β_{p1} in first SAX clustering is determined by the magnitude of amplitude in the whole data set to hierarchically differentiate and obtain the load status in the same magnitude order. And then optimization of clustering results using SAX clustering (second cluster) is performed in load status of same magnitude. Figure 1 (b) and (c) shows the clustering results with 2-SAX algorithm for symbol "a" and data of the heat pump can be represented as a discrete string "bddddcaabbcccccaaabb" with four symbols. Compared with the original 120 sampling points, the load data have realized a great dimensionality reduction and compression. The amplitude breakpoint β_{p2} in the second clustering have a significant effect on the number of load status and the dissimilarity between transformed symbolic string by 2-SAX and original load profiles is gradually reduced by increasing the number of states [32]. However, many more states may result in meaningless of transition probability matrix and big size. Therefore, the number of states is a trade-off between information loss and data compression ratio.

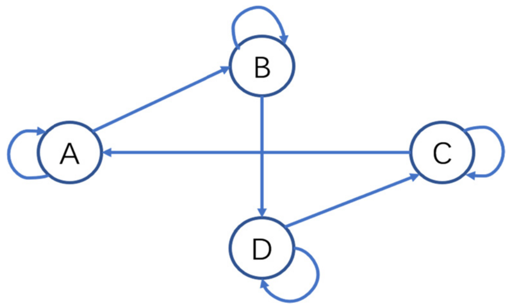

FIGURE 2. **State transition diagram for heat pump.**

C. LOAD PROFILING FOR FEATURES EXTRACTION

The proposed 2-SAX algorithm not only optimise the clustering result of the load states, but, more importantly, provides an opportunity to carry out more accurate load profiling for features extraction. We enquiry the state's transition behavior characteristics and the operation probability characteristics of electrical equipment in this paper.

1) State transition behavior characteristics

Most household appliances are finite-state machines, which have limited on-states and one off-state. The possible state transition behavior can be achieved by analyzing the symbolic string clustered by 2-SAX. Specifically, data of the heat pump in Figure 1 can be represented as a discrete string "bddddcaabbcccccaaabb" and according to that, four states of the heat pump can be self-maintained and state A cannot be converted to C or D directly. Thus, if heat pump was in state A at time k_{i-1}, then it can only be in A or B at time k_i. The same type of characteristics can also be applied backward, for example, the only way to get to C is either from C or D. The state transition diagram (STD) for heat pump has shown in Figure 2. It's worth noting that the state transition behavior is an inherent physical characteristic of devices and will not change due to human operation habits

2) Operation probability characteristics

Specific equipment has higher operation probability in a specific period of one day, depending on its function and costumers' usage patterns. This paper partitions one day into seven fragments: late night (0:00-6:00), early morning (6:00-9:00), morning (9:00-11:00), noon (11:00-2:00), afternoon (2:00-5:00), evening (5:00-8:00), and night (8:00-12:00). In terms of the 2-SAX clustering results, the probability characteristics $P_{AS_p}(t)$ of appliance A operated in state S_p at time t can be conveniently and intuitively calculated, as shown in Equation (3):

$$P_{AS_i}(t) = \frac{N_{AtS_p}}{N_{At}}, \tag{3}$$

where N_{AtS_p} and N_{At} denote the number of state S_p and all states at time t, separately. The operation probability characteristics belong to the costumers' behavior characteristics, which may vary for different resident.

D. MODIFIED INTEGER PROGRAMMING USING APPLIANCE CHARACTERISTICS EXTRACTED BY 2-SAX

An IP-based disaggregation approach is proposed by Suzuki *et al.* [20] and the basic idea of it is to formulate disaggregation as an integer programming problem. Consider a household with N appliances, where the nth appliance ($n = 1, 2, \cdots, N$) has P states. The vector $r = [r_1, r_2, \cdots r_N]$, where $r_n \in R^P$ contains the voltampere (VA) ratings of P states of the nth appliance, can be constructed for all N appliances. The indicator of each state at k time instant is stored in a vector b_k as:

$$b_k[l] \in \{1, 0\} \; for \quad l = 1, 2, \cdots L, L = \sum\nolimits_{n=1}^{N} P, \tag{4}$$

where 1 means that the corresponding state is active while 0 means inactive. Therefore, the total VA reading z_k should be the sum of s_k,VA draws by appliances that is turned ON at that time, which can be expressed as:

$$s_k = Fdiag(b_k)r, \tag{5}$$

$$z_k = hs_k, \tag{6}$$

where $h = [1, 1, \cdots, 1]$, $F = diag[1_1^t, 1_2^t, \cdots 1_N^t]$.

NILM is aimed to find appliance states that are active at time k, namely, to obtain b_k in Equation (5) by using the known quantities z_k, F, r. Hence, the disaggregation problem can be addressed by integer programming, i.e.,

$$\min_{b_k} (z_k - hFdiag(b_k)r)^2. \tag{7}$$

The IP-based disaggregation method may lead to different solutions in different runs even on the same data and yield unsatisfactory results since the infrequently undetected state with high rating [21]. This paper introduces the characteristics of appliance load extracted by 2-SAX previously to modify the IP-based disaggregation solution. Specifically, at first, proposed MIP-AC2S methodology use 2-SAX algorithm to make a dimensional reduction for active power P and apparent power P_s of each device. Vector r_p and r_s derived from the steady-state ratings of each state are determined from the cluster centers of 2-SAX separately. Then, the optimal solution of the Equation (7) is addressed by branch and bound method and modified by load behavior characteristics. The modification can be divided into two stages. The initial stage conducts an output correction of IP solver. If the transitions and relations between consumption behaviors, or rather appliance states, in adjacent periods happens to violate the transition behavior characteristics, the solution must be incorrect and can therefore correct depending on which of the possible states yields lower cost in Equation (7). Operation probability characteristics are applied to correct the ramifications between the disaggregation result of active power and apparent power in second stage. If the disaggregation results of P and P_s are same, the solution is the final result. Otherwise, the total operation probability is calculated for each solution of integer programming respectively, and the highest one is adopted as the ultimate result.

The proposed MIP-AC2S approach for Non-intrusive load disaggregation is illustrated in Figure 3. Using features extracted by 2-SAX not only overcome the shortcomings of IP-based disaggregation approach, also realize a great dimensionality reduction and compression on load data. However, with the size of the dataset rise to "big data", the amount of transmission and computation is still very large. Section 3 designs a Cloud, Edge and End-user synchronizing computing architecture to address the "data deluge" challenge.

III. SYNCHRONIZATION OF CLOUD EDGE AND END-USER COMPUTING

A. CHALLENGE CAUSED BY THE DELUGE OF DATA

With the rapid development of smart grid, AMI has been popularized in many countries, which open opportunities to gather the fine-grained load data. The avalanche of electricity consumption data may cause tremendous pressure on the communication link and data storage computing power of smart grid. Electricity consumption data are collected and distributed on different sites and stored on different substations they belong to. It is costly and time consuming to transmit whole data from each allocated site to cloud data center. Evidently, the pressure exerted by communication, calculation, and storage gradually increases along the data link from local meters to the cloud computing center, as showed in Figure.4.

Most power supply companies now alleviate the burden by reducing the sampling frequency from 15 minutes to a few hours or even days [33]. But the interaction between future grid and user take precision and rapidity as the main direction, which needs high-frequency data for analysis. It is obvious that simply reduce the frequency may lose a lot of effective information and cannot meet the interaction requirements. In addition, a majority of smart meter data analytics methods that are applicable to small data sets may be inappropriate for large data sets [23].

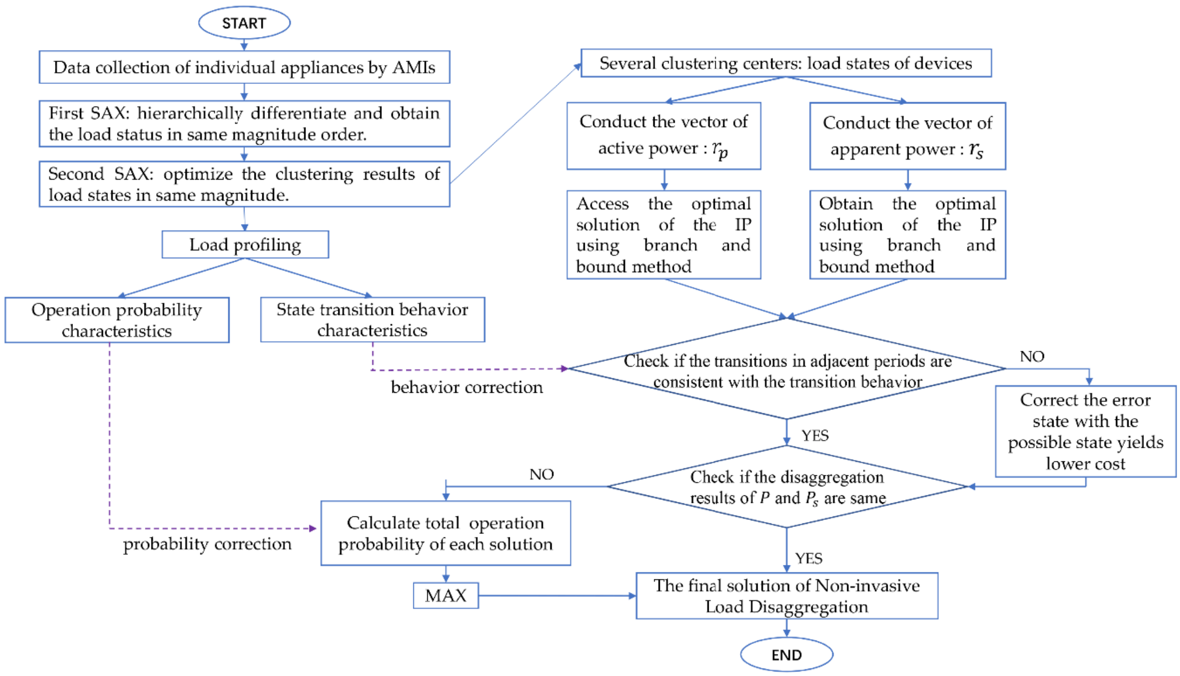

FIGURE 3. Flow diagram of MIP-AC2S algorithm.

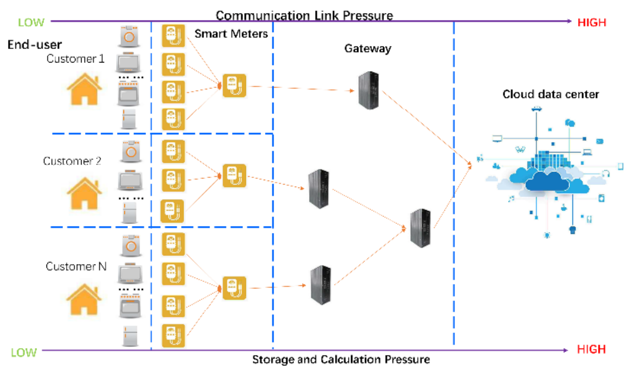

FIGURE 4. The data deluge issue caused by AMI.

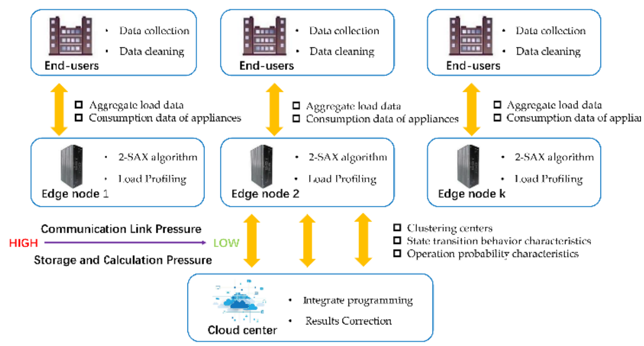

FIGURE 5. Architecture of Cloud, Edge, and End-user synchronizing computing.

B. FRAMEWORK OF CLOUD, EDGE AND END-USER SYNCHRONIZING COMPUTING

Edge computing is a concept of near computing; that is, the operation is completed in the Local Network which is closer to the data source. The cloud computing center will play a role as a central coordinator or a manager, where data that requires further analysis or long-term access instead of all collected data transfer back for processing or storage. Extending the computation to the edge, or even the End-user side, can effectively cut the cost of data storage, communication and processing, remove unnecessary data noise and weaken the impact of transmission delay on data analysis. More data analysis and computing have been transferred back to the edge now.

This paper tries to address the ''data deluge'' issue by introducing a Cloud, Edge and End-user synchronizing computing architecture as showed in Figure 5. The advantage of the proposed architecture is manifested in the following aspects. 1) Applying 2-SAX to transform the load curves into a symbolic string to reduce the data dimensional, considerably, which will substantially contribute to alleviating communication and storage burden. 2) MIP- AC2S methodology is integrated into a divide-and-conquer approach to further improve the efficiency of data processing. Gateways with analysis and calculation capabilities undertake the data compression and load profiling task while integrate programming and corrections are performed at the Cloud data center. 3) This scheme can be performed by directly adding an intelligent gateway with data computing capability or installing a computing module on existing smart meters. It does not substantially change the existing electricity distribution network and considerably reduces procurement and construction costs.

In such case, transmission pressure on the data link and cloud computing center will be substantially alleviated and the proposed MIP-AC2S approach could be further applied toward big data applications.

IV. RESULTS AND DISCUSSION

The data used to verify the validity of the proposed method are obtained from AMPds dataset. AMPds dataset comprehensively records the minutely current reading of a Canadian household and its 19 appliances for 730 days [34], whose sampling frequency is consistent with the existing smart meters for obtaining more realistic results. Five representative household appliances, including heat pump (HPE), cloth dryer (CDE), wall oven (WOE), furnace (FRE) and fridge (FGE), obtained from submetering between August 31 2012, and February 27 2013 is used in our experiment.

To assess the performance of proposed 2-SAX method in data dimension reduction, four widely used criteria are employed, including Compress Ratio (CR), Mean Absolute Percent Error (MAPE), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE):

$$MAPE = \frac{1}{M}\sum\nolimits_{j=1}^{M}\left|\frac{\hat{x}_j - x_j}{\hat{x}_j}\right| *100\% \tag{8}$$

$$MAE = \frac{1}{M}\sum\nolimits_{j=1}^{M}\left|x_j - \hat{x}_j\right|, \tag{9}$$

$$RMSE = \sqrt{\frac{1}{M}\sum\nolimits_{j=1}^{M}(x_j - \hat{x}_j)^2}, . \tag{10}$$

$$CR = \frac{H}{M}, \tag{11}$$

where x_j denotes the actual consumption data and $\hat{x}_j$ denotes the approximative values clustering obtained by 2-SAX.

Two accuracy measures are introduced for performance comparison among our MIP-AC2S method, the traditional IP-based disaggregation and ALIP method proposed in [21].

$$\mathrm{AC}_n = 1 - \frac{\sum_{k=1}^{H}\left|s_k[n] - s_k\hat{[n]}\right|}{2\sum_{k=1}^{H}|s_k[n]|}, \tag{12}$$

$$\mathrm{ACC} = 1 - \frac{\sum_{k=1}^{H}\sum_{n=1}^{N}\left|s_k[n] - s_k\hat{[n]}\right|}{2\sum_{k=1}^{H}\sum_{n=1}^{N}|s_k[n]|}, \tag{13}$$

where $s_k[n]$ is the ground-truth rating of the nth appliance at time index k from the data set, and $s_k\hat{[n]}$ is its estimate obtained by disaggregation.

A. EFFECT OF 2-SAX

Using 2-SAX as the data compression method instead of traditional SAX intends to make the discrete string more consistent with the original consumption status of each devices and to ensure sparsity. Since accurate load profiling requires at least three weeks of symbolical data, the training sample is comprised of the consumption data from five representative household appliances in the first three weeks between August 31 and September 20, 2012. To achieve a fair assessment, data dimensional reduction with SAX and 2-SAX is performed 100 times.

TABLE 1. Equipment status table.

Appliances	States	active power *P* (W)	apparent power P_s (VA)
HPE	HPE_A	5	25
	HPE_B	38	51
	HPE_C	1800	1823
	HPE_D	2400	2450
CDE	CDE_A	0	0
	CDE_B	230	480
	CDE_C	4450	4600
WOE	WOE_A	0	2
	WOE_B	131	180
	WOE_C	2600	2800
	WOE_D	3320	3450
FRE	FRE_A	106	154
	FRE_B	115	167
FGE	FGE_A	0	10
	FGE_B	125	130
	FGE_C	430	445
	FGE_D	1250	1300

Table 1 is the equipment status table obtained by 2-SAX. As shown in Table 1, there is a large order of magnitude difference between state A\B and C\D of HPE and that makes state A and B more likely to be clustered into one state when implementing SAX, while both A and B are crucial and common states of HPE according to the real load data. The same is true for WOE and FGE. This problem can be addressed by proposed 2-SAX and Figure 6 shows the reconstruction performance comparison between SAX (yellow) and 2-SAX (blue). Notably, three states of CDE are distributed in different orders of magnitude and differ greatly from each other, and two states of FRE are in the same order of magnitude, in which 2-SAX has the identical clustering effect as SAX. However, as illustrated, 2-SAX maintains a significant improvement in RMSE, MAE and MAPE in additional three test appliances. In particular, proposed method results in 38.82%, 52.46%, and 13.41% reduction in MAE, MAPE, and RMSE on HPE, respectively, compared with normal SAX with CR=0.5.

It's also illustrated in Figure 6 that the smaller the CR, the better the compression performance is. But the RMSE and MAE will increase in the case of decreasing the compression ratio, which may cause the valuable information missing.

B. LOAD PROFILING FOR FEATURES EXTRACTION

After obtaining the symbolic approximative representation of each household appliances within three weeks, the state

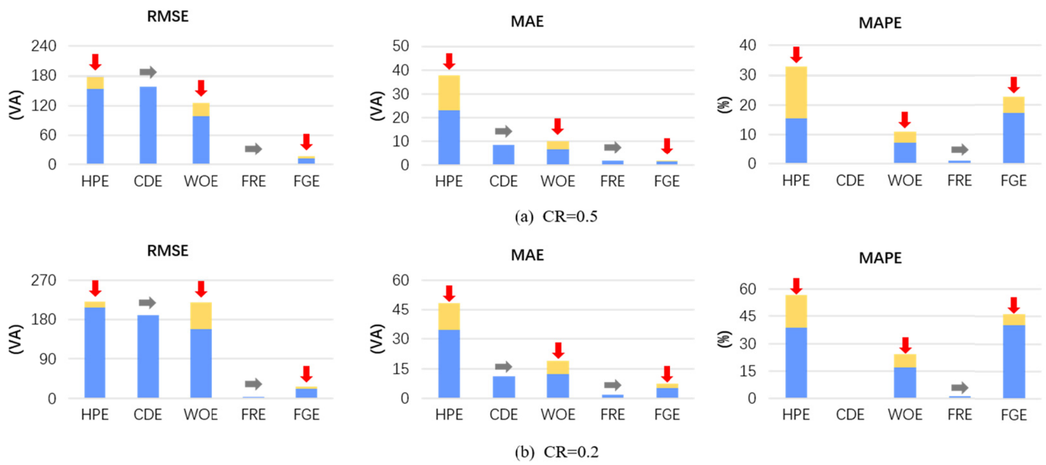

FIGURE 6. Reconstruction performance comparison between SAX (yellow) and 2-SAX (blue). (a) with CR=0.5; (b) with CR=0.2.

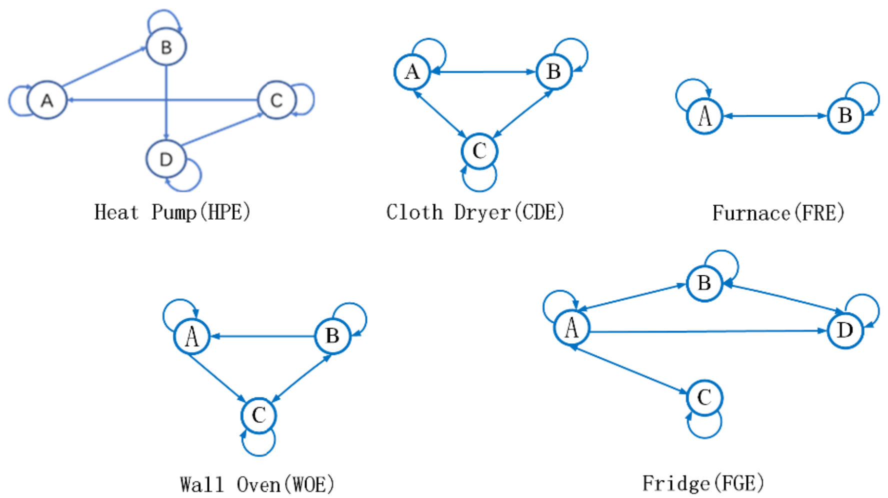

FIGURE 7. Equipment state transition diagram.

TABLE 2. Equipment state operation probability characteristics.

States \ Time	late night	early morning	morning	noon	afternoon	evening	Night
HPE_A	0.09	0.14	0.08	0.03	0.04	0.02	0.04
HPE_B	0.08	0.18	0.15	0.27	0.26	0.25	0.23
HPE_C	0.15	0.05	0.06	0.03	0.03	0.01	0
HPE_D	0	0	0	0	0	0	0.05
CDE_A	0	0	0	0	0	0.02	0.02
CDE_B	0	0	0	0	0.01	0.05	0.04
WOE_A	0	0	0	0	0	0.02	0
WOE_B	0	0	0	0	0	0.01	0.01
WOE_C	0	0	0	0	0	0.01	0.01
FRE_A	0.17	0.14	0.23	0.3	0.3	0.25	0.25
FRE_B	0.16	0.19	0.06	0.04	0.03	0.01	0.04
FGE_A	0.24	0.21	0.2	0.2	0.23	0.20	0.2
FGE_B	0.10	0.09	0.22	0.12	0.1	0.15	0.11
FGE_C	0.01	0	0	0.01	0	0	0
FGE_D	0	0	0	0	0	0	0

transition behavior characteristics and operation probability characteristics could be analyzed and shown in Figure 7 and Table 2.

As showed in Table 2, appliances are differed in operation probability during the same time period of one day, depending on its function and costumers' usage patterns.

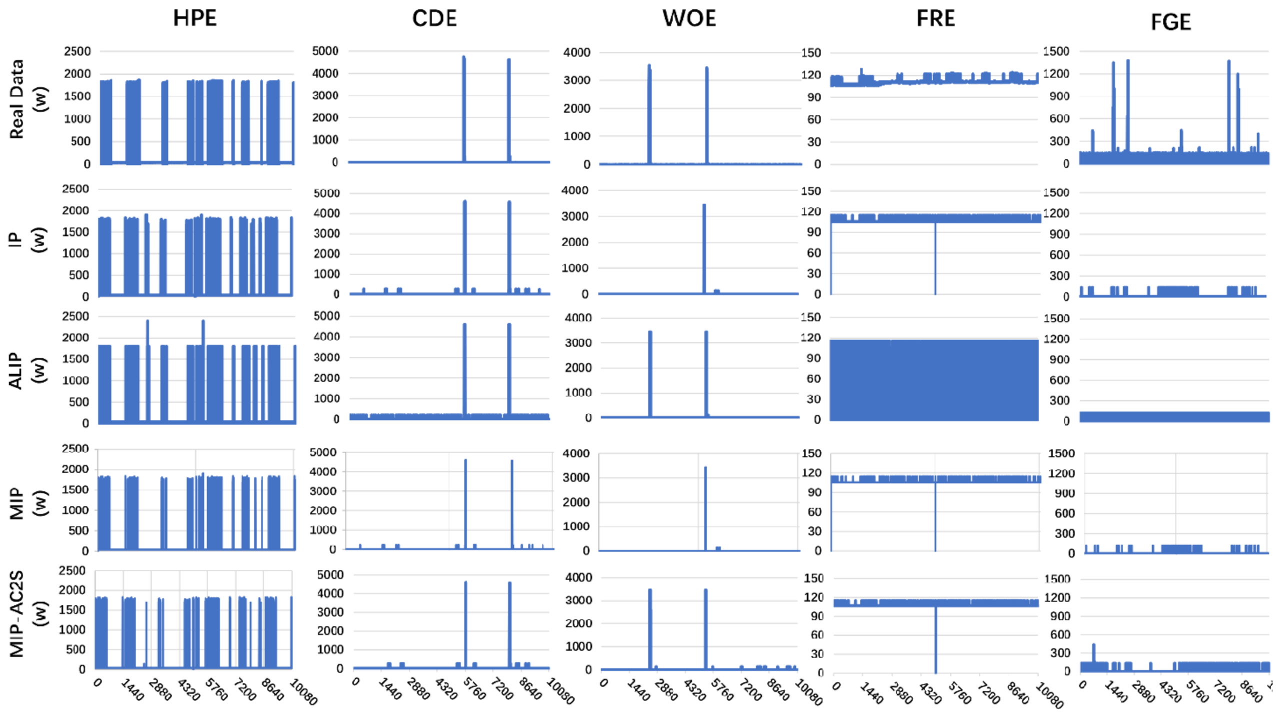

FIGURE 8. Real data and disaggregated appliances' consumption using IP, ALIP, MIP, and MIP-AC2S methods during one week (10080 minutes).

TABLE 3. Disaggregates performance comparison.

Algorithm	AC_HPE	AC_CDE	AC_WOE	AC_FRE	AC_FGE	ACC
IP	0.9608	0.6447	0.6737	0.9777	0.5588	0.8869
ALIP	0.9382	0.7403	0.8868	0.8307	0.7229	0.8730
MIP	0.9628	0.6875	0.6579	0.9778	0.5931	0.8942
MIP-AC2S	0.9746	0.8365	0.9319	0.9779	0.7539	0.9380

The cloth dryer (CDE) and wall oven (WOE) mainly used in evening or night and hardly operated in other periods which are consistent with real data shown in Line 1 from Figure 8. And it's worth to note that operation probability characteristics of FGE_D is zero in all periods cause the spikes are occasional and only occurs 5 times in 10080 sample points.

C. BENCHMARKING OF IP-BASED NILM METHODS

Figure 8 shows the real load consumption data (first line) and the estimated energy consumption contributions for each device using IP (second line), ALIP (third line), MIP (fourth line) and MIPAC2S (last line), respectively. The results illustrate that in a random week chosen for testing, some appliances (like FRE) are fluctuating in a narrow margin during most period, in which ALIP may over-filter the impulsive reading and desegregate a totally wrong solution. In that case, IP, MIP and our MIP-AC2S method perform better, however, our method makes a significant enhancement in "spikes" disaggregation. As for the state FGE_D with 1250 active power, it only occurs 5 times in 10080 sample points and hard to estimate in all methodology.

In test case shown in Table 3, the training sample used to verify the disaggregation performance is comprised of the consumption data from five representative household appliances between August 31 2012, and February 27 2013. The number of states considered for HPE, CDE, WOE, FRE, FGE appliances are 4,3,4,2, and 4, respectively and the total number of samples considered was 10080 × 5 at 1-min intervals. AC and ACC values are given in Table 3. From Figure 8 and Table 3, we observe that MIP-AC2S achieves the best disaggregation performance by having the closest estimate values to the ground truth. Proposed MIP-AC2S method scores 0.2585 better on WOE in AC against IP, 0.1472 better on FRE in AC against ALIP and 0.1608 better on FGE in AC against MIP. The overall outperforms IP, ALIP and MIP significantly.

V. CONCLUSION

A modified IP-based non-intrusive load disaggregation approach is proposed to solve the load disaggregation problem, by adding some corrections based on appliances characteristics extracted by 2-SAX. The 2-SAX instead of normal SAX is implemented in the data compression stage to obtain an approximate discrete representation of original load data for load profiling and state vector construction. States transition behavior characteristics and the operation probability characteristics of electrical equipment extracted by 2-SAX are used to modify the optimal solution of integrate programming problem for overcoming the shortcomings of the previous IP-based approach.

Experimental results demonstrate that 2-SAX results in 38.82%, 52.46%, and 13.41% reduction in MAE, MAPE, and RMSE on HPE and achieves similar performance on the other appliances, compared with normal SAX. Meanwhile, the proposed method MIP-AC2S has a significant accuracy advantage and competitive performance over IP, ALIP and MIP disaggregation method.

Future work can focus on the other modifications on IP-based disaggregation approach and further utilize the potential of the proposed method by adding other external factors, such as temperature, relative humidity, and date, on load profiling.

REFERENCES

[1] (2014). *Grid Modernization and the Smart Grid*. [Online]. Available: http://energy.gov/oe/technology-development/smart-grid

[2] J. Peppanen, M. J. Reno, M. Thakkar, S. Grijalva, and R. G. Harley, "Leveraging AMI data for distribution system model calibration and situational awareness," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 2050–2059, Jul. 2015.

[3] Y. Kwak, J. Hwang, and T. Lee, "Load disaggregation via pattern recognition: A feasibility study of a novel method in residential building," *Energies*, vol. 11, no. 4, p. 1008, 2018.

[4] A. Rahimpour, H. Qi, D. Fugate, and T. Kuruganti, "Non-intrusive energy disaggregation using non-negative matrix factorization with sum-to-k constraint," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4430–4441, Nov. 2017.

[5] R. Godina, E. M. G. Rodrigues, E. Pouresmaeil, J. C. O. Matias, and J. P. S. Catalão, "Model predictive control home energy management and optimization strategy with demand response," *Appl. Sci.*, vol. 8, no. 3, p. 408, 2018.

[6] J. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 420–430, Jan. 2014.

[7] N. Mahmoudi-Kohan, M. P. Moghaddam, M. K. Sheikh-El-Eslami, and E. Shayesteh, "A three-stage strategy for optimal price offering by a retailer based on clustering techniques," *Int. J. Electr. Power Energy Syst.*, vol. 32, no. 10, pp. 1135–1142, 2010.

[8] J. D. Rhodes, W. J. Cole, C. R. Upshaw, T. F. Edgar, and M. E. Webber, "Clustering analysis of residential electricity demand profiles," *Appl. Energy*, vol. 135, pp. 461–471, Dec. 014.

[9] G. Sun, Y. Cong, D. Hou, H. Fan, X. Xu, and H. Yu, "Joint household characteristic prediction via smart meter data," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1834–1844, Mar. 2019.

[10] H.-H. Chang, K.-L. Lian, Y.-C. Su, and W.-J. Lee, "Power-spectrum-based wavelet transform for nonintrusive demand monitoring and load identification," *IEEE Trans. Ind. Appl.*, vol. 50, no. 3, pp. 2081–2089, May/Jun. 2014.

[11] T. Hassan, F. Javed, and N. Arshad, "An empirical investigation of V-I trajectory based load signatures for non-intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 870–878, Mar. 2014.

[12] S. Kong, Y. Kim, R. Ko, and S.-K. Joo, "Home appliance load disaggregation using cepstrum-smoothing-based method," *IEEE Trans. Consum. Electron.*, vol. 61, no. 1, pp. 24–30, Feb. 2015.

[13] S. M. Tabatabaei, S. Dick, and W. Xu, "Toward non-intrusive load monitoring via multi-label classification," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 26–40, Jan. 2017.

[14] S. Makonin, F. Popowich, I. V. Bajic, B. Gill, and L. Bartram, "Exploiting HMM sparsity to perform online real-time nonintrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2575–2585, Nov. 2016.

[15] J. A. Mueller, A. Sankara, J. W. Kimball, and B. McMillin, "Hidden Markov models for nonintrusive appliance load monitoring," in *Proc. North Amer. Power Symp.*, Sep. 2014, pp. 1–6.

[16] J. Kelly and W. Knottenbelt, "Neural NILM: Deep neural networks applied to energy disaggregation," in *Proc. 2nd ACM Int. Conf. Embedded Syst. Energy-Efficient Built Environ.*, Nov. 2015, vol. 206, no. 3, pp. 55–64.

[17] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton, "Sequence-to-point learning with neural networks for non-intrusive load monitoring," in *Proc. 32nd AAAI Conf. Artif. Intell.*, New Orleans, LA, USA, Feb. 2018, pp. 2604–2611.

[18] M. Matsumoto, Y. Fujimoto, and Y. Hayashi, "Energy disaggregation based on semi-binary NMF," in *Proc. Int. Conf. Mach. Learn. Data Mining Pattern Recognit.*, vol. 9729, 2016, pp. 401–414.

[19] S. Singh and A. Majumdar, "Deep sparse coding for non–intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4669–4678, Sep. 2017.

[20] K. Suzuki, S. Inagaki, T. Suzuki, H. Nakamura, and K. Ito, "Nonintrusive appliance load monitoring based on integer programming," in *Proc. SICE Annu. Conf.*, Aug. 2008, pp. 2742–2747.

[21] M. Z. A. Bhotto, S. Makonin, and I. V. Bajić, "Load disaggregation based on aided linear integer programming," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 64, no. 7, pp. 792–796, Jul. 2017.

[22] M. A. Zehir, A. Batman, and M. Bagriyanik, "Review and comparison of demand response options for more effective use of renewable energy at consumer level," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 631–642, Apr. 2016.

[23] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Trans. Smart Grid*, to be published.

[24] X. Tong, C. Kang, and Q. Xia, "Smart metering load data compression based on load feature identification," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2414–2422, Sep. 2016.

[25] M. P. Tcheou *et al.*, "The compression of electric signal waveforms for smart grids: State of the art and future trends," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 291–302, Jan. 2014.

[26] R. Mehra, N. Bhatt, F. Kazi, and N. M. Singh, "Analysis of PCA based compression and denoising of smart grid data under normal and fault conditions," in *Proc. IEEE Int. Conf. Electronics, Comput. Commun. Technol.*, Jan. 2013, pp. 1–6.

[27] J. C. S. De Souza, T. M. L. Assis, and B. C. Pal, "Data compression in smart distribution systems via singular value decomposition," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 275–284, Jan. 2017.

[28] Y. Wang, Q. Chen, C. Kang, Q. Xia, and M. Luo, "Sparse and redundant representation-based smart meter data compression and pattern extraction," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2142–2151, May 2017.

[29] Y. Wang, Q. Chen, C. Kang, and Q. Xia, "Clustering of electricity consumption behavior dynamics toward big data applications," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2437–2447, Sep. 2016.

[30] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proc. 8th ACM SIGMOD Workshop Res. Issues data Mining Knowl. Discovery*, Jul. 2013, pp. 2–11.

[31] J. Sevcech and M. Bielikova, "Symbolic time series representation for stream data processing," in *Proc. IEEE Trustcom/BigDataSE/ISPA*, Helsinki, Finland, Aug. 2015, pp. 217–222.

[32] D. Niu, H. Shi, J. Li, and C. Xu, "Research on power load forecasting based on combined model of Markov and BP neural networks," in *Proc. 8th World Congr. Intell. Control Autom.*, Jinan, China, Jul. 2010, pp. 4372–4375.

[33] A. Unterweger and D. Engel, "Resumable load data compression in smart grids," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 919–929, Mar. 2015.

[34] S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. V. Bajic, "AMPds: A public dataset for load disaggregation and eco-feedback research," in *Proc. IEEE Elect. Power Energy Conf.*, Aug. 2013, pp. 1–6.

YAOXIAN LIU was born in 1990. He received the B.S. degree in communication engineering and the M.S. degree in detection technology and automatic equipment from North China Electric Power University (NCEPU), Beijing, China, in 2013 and 2016, respectively, where he is currently pursuing the Ph.D. degree in electrical engineering. His research interests include non-intrusive load monitoring, demand response, and user behavior.

YI SUN was born in 1972. He received the M.S. degree in communication and information system and the Ph.D. degree in electric information technology from North China Electric Power University (NCEPU), Beijing, China, in 2009 and 2014, respectively, where he is currently a Professor of information and communication engineering. His research interests include smart power consumption, demand response, and power system communication technology.

BIN LI was born in Beijing, China, in 1983. He received the B.S. and Ph.D. degrees from the State Key Laboratory of Information Photonic and Optical Communications (IPOC), Beijing University of Posts and Telecommunications (BUPT), in 2005 and 2010, respectively. He was a joint training Doctor with Yuan Ze University (YZU), Taiwan, focused on the antenna design, antenna measurement, electromagnetic scattering, and asymptotic high frequency techniques.

He was with the Wireless and Optical Networking Research Group, Bell Labs Research, China. He joined North China Electric Power University (NCEPU), in 2011, where he is currently an Associate Professor with the Communication Technology Research Center (CTRC), School of Electric and Electronic Engineering. He has published more than 200 journals and conference papers, and submitted more than 20 patents. His research interests include hybrid optical and wireless communication networks, the next generation Internet, and electric power communication, network routing, and signaling technology.

Dr. Li is a member of the China Communications Standards Association (CCSA) and the Chinese Society for Electrical Engineering (CSEE). He has received the certifications from IBM Rational University Program and always working as an OPNET Training Engineer.