Leveraging Multi-Granularity Energy Data for Accurate Energy Demand Forecast in Smart Grids

Zhichuan Huang and Ting Zhu
University of Maryland, Baltimore County
{zhihu1, zt}@umbc.edu

***Abstract*—Accurate energy demand prediction is very important for smart grids to conduct demand response and stabilize the grids. In previous work, many prediction algorithms are proposed to improve the energy consumption prediction accuracy based on the aggregated energy consumption in the whole grid. Recently, with the increasing installations of smart meters in individual homes, high granularity (e.g., per minute) energy consumption data in individual homes becomes available and provides us a great opportunity for better energy consumption prediction. In this paper, we propose M-Pred to utilize the high granularity energy consumption data collected by smart meters in individual homes for better energy consumption prediction in smart grids. In M-Pred, we propose a learning algorithm to learn energy consumption patterns of individual homes from the high granularity energy consumption data. The consumption patterns we learn from homes are then applied for energy consumption prediction in smart grids. Furthermore, since not every home in a smart grid is equipped with a smart meter, we propose a matching and prediction algorithm to leverage the multi-granularity energy data for accurate consumption prediction. We conducted extensive system evaluations with 726 homes' minute-level power consumption data for more than 12 months. The simulation results show that our design can provide accurate energy consumption prediction for the next hour with negligible errors (e.g., Mean Absolute Percentage Error is 2.12%).**

I. Introduction

Compared to traditional power grid, smart grids i) are expected to be robust against grid disturbance or outage; ii) can use more environmental friendly renewable resources; and iii) can utilize the rich information from homes for better energy management. However, to better utilize energy generation in smart grids, one main challenge is how to accurately predict the energy demand (i.e., consumption). This is because generators in smart grids need to generate enough power for energy usage to avoid power outage. To improve the accuracy of energy demand prediction, some prediction algorithms [5, 9] have been proposed based on the aggregated energy consumption in the whole grid. One major limitation of these approaches is that they are predicting energy consumption for the next day or a even longer time period. To achieve faster demand response in smart grids, energy demand prediction for the immediate near future (e.g., the next hour) is more desirable. Another limitation of the existing approaches is that they only predict the hourly average energy consumption. However, energy consumption over an hour can change dynamically. If generators only generate the energy based on the average energy consumption, it is highly possible that the peak power demand will be higher than the power generated in smart grids. Thus, it is very important to predict not only average hourly energy consumption but also the peak demand within an hour. To address these limitations, we propose to utilize both the energy consumption in individual homes and aggregated energy consumption in smart grids for more accurate energy consumption prediction. With the increasing installations of smart meters in individual homes nowadays, high granularity (e.g., per minute) energy consumption data in individual homes becomes available [20, 14]. Because high granularity energy consumption in individual homes provide more information of consumption patterns, high granularity energy consumption data from some of the individual homes with smart meters provides us a great opportunity for better energy consumption prediction.

To utilize the high granularity energy consumption data from some of individual homes, there are three big data related challenges: i) the huge amount of high granularity energy consumption data collected from homes introduces the severe data storage issue; ii) the energy consumption prediction operation should have low computation overhead for faster demand response; and iii) not every home in the power grid has a smart meter for monitoring high granularity energy consumption data, thus we need to predict energy consumption and peak demands based on different granularities of energy consumption data from different homes. To address these three challenges, we propose M-Pred for accurate energy consumption and peak demand prediction in smart grids. In M-Pred, to reduce the large amount of data storage, we propose to store the energy consumption patterns instead of the whole energy consumption data. Thus, a learning algorithm is proposed to learn energy consumption patterns of individual homes from their energy consumption data. To reduce the computation complexity and communication overhead for energy consumption prediction, we design a distributed energy consumption prediction algorithm to utilize the consumption patterns of homes. Furthermore, considering not every home in power grids has a smart meter, we propose a matching algorithm to cope with different granularity of the collected energy consumption patterns from individual homes and aggregated energy consumption in the grids. Then the matching results can be applied to utilize the energy consumption patterns collected from homes that have smart meters. The main contributions of this paper are as follows:

• To the best of our knowledge, this is the first work to utilize the spatial and temporal features of the detailed power consumption in individual homes for more accurate power consumption prediction in smart grids.

• With the analysis of massive minute-level power consumption data, we propose **M-Pred** to learn energy consumption patterns of individual homes from their energy consumption data and then utilize these patterns to predict the power consumption in smart grids. Considering that not every home in smart grids has a smart meter, we also propose a matching and prediction algorithm for accurate energy consumption prediction in such smart grids.

• To validate our design, we evaluate our work extensively with more than 12 months' empirical minute-level power consumption data from 726 homes'. The evaluation results show that our design can provide accurate energy consumption prediction for the next hour with Mean Absolute Percentage Error (MAPE) of 2.12%.

The rest of the paper is organized as follows: the motivation of this project is introduced in §II; the problem formulation and our detailed design are described in §III and §IV, respectively; implementations and simulations are provided in §V; related work is discussed in §VI; finally, we conclude our paper in §VII.

II. Motivation

In this section, we first explain that why peak demand forecast is needed compared to hourly average power consumption in smart grids. Then we give an example from our empirical data to demonstrate that the power consumption patterns of high time granularity data at individual homes can be beneficial for better prediction. Finally, we summarize the opportunities and challenges of utilizing high time granularity energy consumption data in individual homes.

A. The Need for Peak Demand Forecast

Peak demand is crucial in real-time demand response applications in smart grids. In this section, we investigate the relationship between hourly average power consumption and peak demand during the hour. We collect the minute-level power consumption data from 726 homes within a city. For hourly average power consumption, it is defined as the average minute-level power consumption within one hour and peak demand is defined as the maximum power consumption in each hour. The relationship between hourly average power consumption and peak demand of 726 homes is shown in Figure 1. For the same hourly power consumption in these homes, the peak demand during the hour can be quite diverse. For example, when hourly average power consumption is around $672kW$, the difference between peak demand and hourly average power consumption varies from $1.16MW$ to $1.85MW$, which is almost $700kW$. Furthermore, with the higher hourly average power consumption, the diversity of the peak demand is more significant. This indicates hourly power consumption misses a lot of vital information on how homes consume energy. The missing information can be crucial for energy generation control and scheduling in smart grids. For example, if we only predict the hourly power consumption in future, the generators may generate either too much energy or not enough energy to cause power outage. Therefore, it is important to not only predict the hourly average power consumption but also peak demand within an hour.

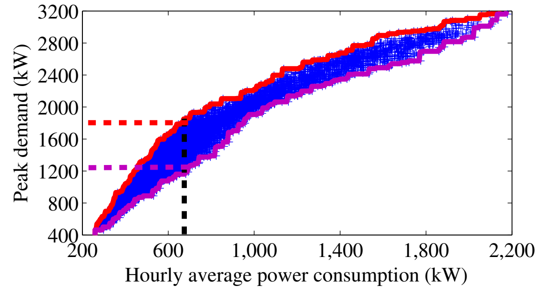

Figure 1: Average VS peak hourly energy consumption

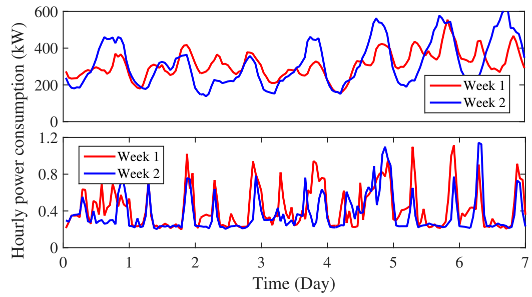

Figure 2: Hourly power consumption over two weeks (the top figure shows the aggregated power consumption in the smart grid, and the bottom figure shows the power consumption data from a single home)

B. The Need for Power Consumption in Individual Homes

One main limitation of existing prediction approaches is that they only utilize the aggregated power consumption in smart grids to predict hourly power consumption in short-term. Unfortunately, because not every home in smart grids has the same consumption pattern, the aggregated power consumption in smart grids does not show strong correlation over time. The aggregated hourly average power consumption in a smart grid over two weeks is shown in top figure of Figure 2. We can find that the power consumption patterns in each day are quite different. Besides, the power consumption pattern in two consecutive weeks are also different. Therefore, if we utilize previous day's power consumption to forecast current day's power consumption with aggregated data, the forecast accuracy would be low. In the meanwhile, from the bottom figure of Figure 2, we

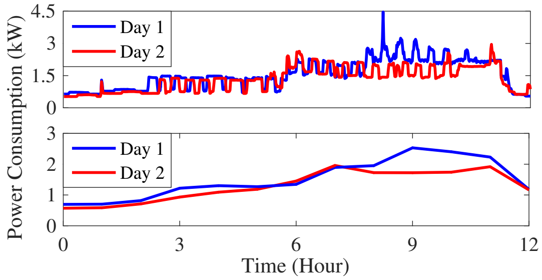

Figure 3: Power consumption pattern in a single home

find that power consumption patterns of a single home in consecutive two weeks are quite similar. The reason is that for a single home, the weekly activities of homeowner are normally fixed with minor varieties. While in a smart grid, the minor varieties of different homes aggregate together, which causes that the aggregated power consumption patterns of different periods are different. Therefore, if we can utilize the power consumption patterns in individual homes, power consumption in smart grids can be better predicted.

C. The Need for High Time Granularity Data

In previous sections, we show hourly energy consumption in smart grids has weak correlation over time. Thus the energy consumption prediction based on aggregated energy consumption is not accurate. In this section, we investigate the correlation of high granularity energy consumption data in a single home. Hourly energy consumption and minute-level energy consumption data of a single home are shown in Figure 3. The top figure is the minute-level energy consumption in a single home for 12 hours. We can find the energy consumption pattern of two days are quite similar except there are several minutes delay between two days. This is because homeowner in a single home has relatively stable behavior pattern, which consumes similar amount of energy. However, the hourly energy consumption does not show the same phenomenon in the bottom figure. This is because even though homeowner has similar energy consumption pattern, the small delay of energy consumption makes the energy consumption pattern disappear in hourly energy consumption. Thus, to realize real-time energy consumption prediction, it is important to utilize the minute-level energy consumption data in individual homes.

D. Challenges of Utilizing Energy Consumption Patterns

From the empirical results of previous sections, we find out energy consumption patterns of homeowner can be explored with high granularity energy consumption data in individual homes. Then the explored energy consumption pattern in individual homes can be beneficial for energy consumption prediction in smart grids. However, there are many challenges for utilizing energy consumption pattern in individual homes. First of all, high granularity energy consumption data that reveals energy consumption pattern are large amount of data especially for large amount of homes in smart grids. Thus, we need an efficient algorithm to first learn energy consumption patterns in individual homes and then use the learned consumption patterns for energy consumption prediction. Secondly, in reality, not every home in the smart grids is deployed with a smart meter for monitoring high granularity energy consumption data. Thus, it is important to investigate how to utilize partial of the high granularity energy data from homes in smart grids to realize real-time energy consumption prediction. To address these challenges, we propose M-Pred, which utilizes different granularity of energy consumption data from individual homes for energy consumption prediction in smart grids.

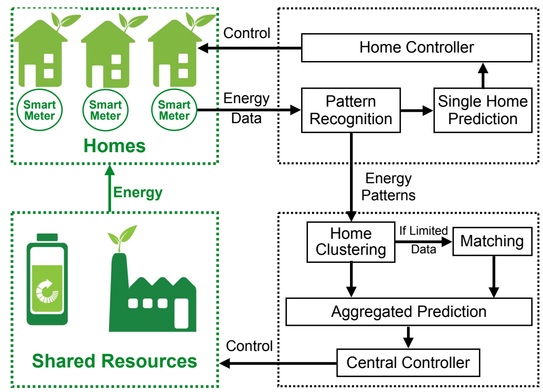

Figure 4: System Overview of M-Pred

III. Problem Formulation

In this section, we provide an overview of M-Pred, describe how energy consumption in individual homes can be utilized and present our design goal.

A. Overview of M-Pred

To realize real-time energy consumption prediction in smart grids, we propose M-Pred, which utilizes the high granularity energy consumption data from individual homes to predict energy consumption in smart grids. The system overview of our design is shown in Figure 4. In smart grids, some homes are deployed with smart meters for monitoring high granularity energy consumption data. The energy consumption data is then processed locally for pattern recognition (detailed in § IV-A). The patterns recognized can be first utilized for energy consumption prediction in a single home. Then, to reduce the computation complexity and communication overhead for energy consumption forecast, a distributed energy consumption prediction algorithm is proposed to utilize the consumption patterns learned from homes. The learned energy consumption patterns from different homes will be clustered and then applied for energy consumption prediction in smart grids (detailed in § IV-B). Considering that not every home in smart grids is deployed with a smart meter, an energy matching algorithm is proposed for smart grids in which only partial homes are deployed with smart meters (detailed in § IV-C). Finally, the energy consumption prediction can be latter used for generation scheduling and control to avoid power outage and improve the energy efficiency in smart grids.

Notations	Definitions
N	Number of home in a smart grid
$p_i(t)$	Original power consumption of home i at t
$d_i(t)$	Predicted power consumption of home i at t
$e_i(t)$	Prediction error of home i at t
$h_i(t)$	Predicted peak demand of home i at t
S	Power consumption pattern set S
$sdist(X_1, X_2)$	Distance between two vectors X_1 and X_2
c_{ij}	Consumption correlation between home i and j

Table I: Definitions of notations

B. Design Goal

To predict power consumption with high time granularity in short term, we propose a middleware M-Pred, which is designed to be run at both local smart meters and central servers. Due to the large amount of high granularity energy consumption data, the computation and implementation complexity of M-Pred should be low. Let $\{p_i(1), \cdots, p_i(t)\}$ be the original minute-level power consumption series for home i, $\{p(1), \cdots, p(t)\}$ be the aggregated energy consumption in smart grids, and $\{\hat{d}(1), \cdots, \hat{d}(m)\}$ and $\{\hat{h}(1), \cdots, \hat{h}(m)\}$ be the peak demand and hourly average power consumption in smart grids, respectively. Then we have

$$h(i) = \sum_{j=1}^{j=N} \sum_{k=i*60-59}^{k=i*60} p_j(k) \tag{1}$$

$$d(i) = \max_{k=i*60-59}^{k=i*60} \sum_{j=1}^{j=N} p_j(k) \tag{2}$$

Our design goal is to minimize the prediction error of both hourly average power consumption $\sum_{i=1}^{i=m} \{h(i) - h(i)\}$ and peak demand $\sum_{i=1}^{i=m} \{d(i) - d(i)\}$, where $h(i)$ and $d(i)$ are predicted hourly average and peak demand. Note that our proposed design M-Pred is used to provide accurate power consumption (hourly average and peak demand) prediction in both a single home and smart grids. Also, due to different applications, the energy power prediction should also be able to conducted for different prediction window size (e.g., next 1 hour, 4 hours or 1 day).

IV. System Design

In this section, we introduce the main design of M-Pred. Our design consists of three parts: i) power consumption patterns recognition in a single home; ii) power consumption prediction in smart grids; iii) power consumption prediction with limited data from individual homes. In the end, we analyze the performance and complexity of M-Pred.

A. Consumption Pattern Recognition

Different from previous prediction model based on historical data, we need to predict the power consumption in the immediate future (e.g., next minute) for real-time control because of the limited energy storage units in smart grids. Thus the power consumption data of yesterday or last month can be much less useful. And because there are different power consumption signatures for different loads, we can predict the data based on the detected power consumption signatures. To evaluate our idea, we use trace data of one year for one home to investigate the correlation between power consumption of different time gaps.

Algorithm 1 Pattern Recognition Algorithm

```
S = ∅;
while t < T do
    ρ_max = 0, index = −1;
    for detected consumption pattern S_i do
        Calculate ρ_i(t) based on Equation (3);
        if ρ_i(t) > ρ_max then
            ρ_max = ρ_i(t), index = i,
        end if
    end for
    if ρ_max > ρ_min then
        Detect a new consumption pattern S_new and add to S;
        t = t + length(S_i);
    else
        t = t + 1;
    end if
end while
```

We first run a power consumption pattern detection algorithm on the data set to recognize the power consumption patterns. In this paper, we use a Euclidean distance-based function to quantify the similarity between two vectors. The distance between two vectors can be calculated as:

$$\rho_{i,j} = \frac{1}{l(S_i)} \sum_{t=1}^{l(S_i)} (S_i(t) - S_j(t))^2 \tag{3}$$

If the distance of two vectors calculated through Equation (3) is small, then the similarity of two vectors is high. Then we go through the whole data set to find the possible consumption patterns. To simplify the algorithm, we use fixed length of energy consumption patterns. The algorithm we use is shown in Algorithm 1. At the beginning, the consumption pattern set S is empty. For $t < T$, we calculate similarity between power consumption data and recognized consumption patterns based on Equation (3). If we find the similarity between current power consumption and existing power consumption pattern is higher than current maximum similarity, we reassign maximum similarity and mark $index = i$. Then we compare the maximum similarity we find to the threshold of minimum similarity ρ_{min}. If $\rho_{max} > \rho_{min}$, we then detect a new consumption pattern S_{new} and add it to consumption pattern set S. Then, we update $t = t + l(S_{new})$ for further recognition. Otherwise, we update $t = t + 1$ to continue the recognition process.

To evaluate the performance of our consumption pattern recognition algorithm, we show some of the recognized

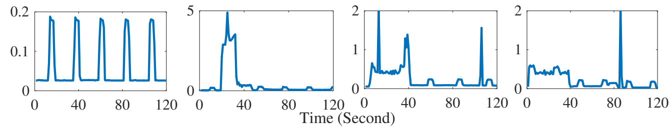

Figure 5: Some examples of signatures

consumption patterns from one home in Figure 5. For example, the left top figure is the periodical activity of refrigerators while other three consumption patterns are the combinations of several appliances activities.

1) Consumption Pattern Fitting: With the recognized consumption patterns in individual homes, future power consumption can be predicted. However, to enable power consumption prediction, the length of energy consumption pattern have to be long enough (e.g., 120 minutes in Figure 5). To save the storage in smart meter and communication overhead for future energy consumption prediction in central server, we introduce polynomial curve fitting to sketch the power consumption patterns. To fit power consumption patterns, we first consider the general form for a polynomial of degree n:

$$p(t) = a_0 + a_1 x(t) + a_2 x(t)^2 + a_3 x(t)^3 + \cdots + a_n x(t)^n \quad (4)$$

The curve that gives minimum error between real power consumption pattern and the fitted curve is best. In our case, we use least squares error to find the best fitted curve of power consumption patterns. The general expression for any error using the least squares approach is:

$$err = \sum_{t=1}^{T_p} (d(t) - p(t))^2 \quad (5)$$

We then find the A to minimize err where A is $[a_0, a_1, \cdots, a_n]^T$. To minimize Equation (5), take the derivative with respect to each coefficient set to zero:

$$\frac{\partial err}{\partial a_j} = -2 \sum_{t=1}^{T_p} (d(t) - p(t)) x(t)^j \quad (6)$$

Then we have to solve $n+1$ equations to find A to minimize err:

$$XA = B \quad (7)$$

where

$$X = \begin{bmatrix} 1 & \sum x(t) & \sum x(t)^2 & \dots & \sum x(t)^n \\ \sum x(t) & \sum x(t)^2 & \sum x(t)^3 & \dots & \sum x(t)^{n+1} \\ \sum x(t)^2 & \sum x(t)^3 & \sum x(t)^4 & \dots & \sum x(t)^{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x(t)^n & \sum x(t)^{n+1} & \sum x(t)^{n+2} & \dots & \sum x(t)^{n+n} \end{bmatrix} \quad (8)$$

$$B = \begin{bmatrix} \sum d(t) \\ \sum x(t) d(t) \\ \sum x(t)^2 d(t) \\ \dots \\ \sum x(t)^n d(t) \end{bmatrix} \quad (9)$$

Then we can get A from $X^{-1}B$. With A, we use Equation (4) to calculate fitted curve. After fitting, we only need to store A instead of $\{p(1), \cdots, p(t)\}$ for future power consumption prediction.

2) Power Consumption Prediction in A Single Home: With the fitted power consumption pattern A, we can easily recover the power consumption pattern with Equation 7. With the recognized power consumption patterns and historical data, we can predict the energy consumption in future. The process is similar to pattern recognition. For each recognized power consumption pattern S_i, we calculate the similarity between power consumption pattern and historical data $P = \{p(1), \cdots, p(t)\}$ based on Equation 3. Then based on the distance between historical data and consumption pattern, we predict the future power consumption as

$$d(t+k) = s_i(t+k) + sdist(S_i, P) \quad (10)$$

$sdist(S_i, P)$ is the Euclidean distance between two vectors S_i and P, which can be calculated with Equation 3.

B. Power Consumption Prediction in Smart Grids

In previous section, we investigate the power consumption patterns in individual homes. In this section, we describe how we can utilize the recognized consumption patterns to predict future aggregated power consumption in smart grids for real-time demand response.

1) Individual Homes Clustering: A simple solution to utilize recognized consumption pattern for aggregated power consumption prediction is to conduct predictions in each home and then send the prediction results from each smart meters to the central server. However, with large number of homes in smart grids, it is not scalable because this solution would introduce high storage need and communication overhead between smart meters and central server. Because homes in the same area may have the similar power consumption patterns, in this section, we describe how to use the spatial correlation among power consumption of homes for power consumption prediction. However, different homes will not have different correlations at different time. Thus, we need to keep updating the correlations among homes for prediction. To evaluate our idea, we use trace data of 726 homes for one month to investigate the correlation among power consumption of different homes. The spatial correlation between 726 homes and one single home i is shown in Figure 6. X-axis is the standard correlation between two vectors. The correlation between two homes can be calculated as:

$$c_{ij}(t) = \frac{1}{l} \sum_{k=t-l}^{t} (p_i(k) - p_j(k) - \frac{1}{l} \sum_{k=t-l}^{t} (p_i(k) - p_j(k)))^2 \quad (11)$$

Y-axis is the CDF of given correlation. In Figure 6, most of the homes have similar consumption patterns with home i and 20% of the homes have correlation less 0.1 with home i.

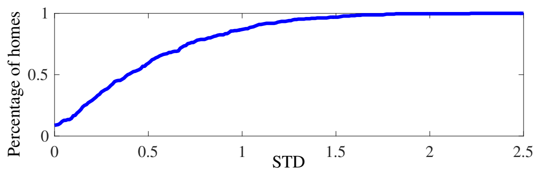

Figure 6: Spatial correlation of different homes over time

Algorithm 2 Consumption Pattern Updating Algorithm

Input: Fitting errors $e(t)$ between real secondly power consumption and polynomial fitting curve $p(t) - d(t)$.
Output: Consumption pattern set S.

```
1:  for t = 1 to T do
2:    if |e(t)| > d_h then
3:      power = e(t), start = t, i = t;
4:      while i < T & |e(i)| > d_h do
5:        i++;
6:      end while
7:      end = i, Add {power, start, end} to P;
8:    else
9:      for j = 1 to t_p do
10:       if |e(t) − e(t + j)| ≤ θ then
11:         per = j;
12:         for k = 1 to t_k do
13:           if |e(t + k) − e(t + k + j)| ≥ θ then
14:             time = k;
15:           end if
16:         end for
17:         Add {power, start, end, per, time} to S;
18:       end if
19:     end for
20:   end if
21: end for
```

Thus, it is possible to infer power consumption of multiple homes based on the prediction results of a single homes.

2) Updating Power Consumption Patterns: Based on analysis of § IV-A, we show how to generate consumption pattern from real power consumption and polynomial fitting results. The key idea is to discover valuable power consumption data by comparing differences between real power consumption and polynomial fitting results. The detail of the algorithm is described in Algorithm 2. For each second t from 1 to T, it checks if the fitting error is larger than threshold of high power consumption d_h (Lines 1-2). Note that d_h needs to be selected carefully. If d_h is too small, it has better performance but cost too much storage; if d_h is too large, the error of polynomial fitting will be too high. Then if $|e(t)| > d_h$, it finds the endtime of high power consumption and add $\{power, start, end\}$ to P (Lines 3-7). If the fitting error is less than the threshold, then the algorithm checks if there is periodical power consumption (Line 8-10). If yes, it finds the period and last time of periodical power consumption (Lines 11-16), then adds $\{power, start, end\}$ to P (Lines 17-21).

Algorithm 3 Correlation of Power Consumption

Input: Fitting errors $e_1(t)$ and $e_2(t)$ of two days.
Output: Decision of correlation of two days.

```
1:  count = 0;
2:  for t = 1 to T do
3:    if |e_1(t) − e_2(t)| < d_c then
4:      count++;
5:    end if
6:  end for
7:  if count ≥ T/2 then
8:    Correlation can be used to update pattern;
9:  else
10:   Correlation can not be used to update pattern;
11: end if
```

3) Utilization of Consumption Correlation: Based on the analysis of § IV-B1, power consumption of different days may have similar patterns and thus only one day's consumption needs to be stored for reducing storage space. For example, instead of storing new power consumption data at Day 3, we only need to store the power consumption differences between Day 1 and Day 3. However, we need to carefully determine whether we can reduce storage space based on the correlation between two days. The format for storing correlation information is as follows: $\{b_c,\ t_1,\ e_1,\ \cdots,\ t_m,\ e_m\}$. b_c is the bit used to store whether or not we utilize the correlation of power consumption in storing the data. t_i ($i \in \{1, \cdots, m\}$) is used to store the time slot that power consumption of two days are different and e_i ($i \in \{1, \cdots, m\}$) is used to store the power consumption differences. As long as $m < T/2$, the data used to store time and power consumption differences ($2 * m$ data points) would be less than directly storing fitting errors (T data points). Algorithm 3 gives a detailed description on how to make the decision. For each second of t from 1 to T, it checks if the difference of power consumption for two days is less than threshold d_c and counts the number of time slot (Lines 1-6). Note that d_c needs to be selected carefully. If d_c is too small, it has better performance but cost more storage to store high power consumption; if d_c is too large, the error of polynomial fitting wil be too high. Then algorithm checks if correlation between two days can be used to save space (Lines 7-11). In server, when power consumption data of a new day is received from a smart meter, it runs Algorithm 3 using the power consumption of past several days to find a day to be used to save space. If it does not find, it stores the power consumption data of that day from smart meter to database.

4) Power Consumption Prediction with Low Computation: Based on the correlation, we can predict $p(t)$ based on readings from other homes:

$$p(t) = \sum_{i=1}^{N} p_i(t) = \sum_{i=1}^{N} \sum_{j=1}^{N_h} \frac{p_j(t) * c_{ij}(t)}{\sum_{j=1}^{N} c_{ij}(t)} \tag{12}$$

If $c_{ij}(t)$ does not exist, then we replace $c_{ij}(t)$ as $c_{ij}(t_k)$ where t_k is the latest time for updating correlation between home i and j. N_h is the number of homes selected for conducting power consumption prediction based on correlation among homes. The prediction accuracy is highly dependent on the selection of homes for power consumption prediction. Here we give a detailed description on how to make the selection. We first add home 1 as one of the selected homes. For each home i from 2 to N, it checks if the correlation of power consumption between home i and any selected homes is smaller than threshold d_c. If the correlation between home i and one of the selected homes is smaller than d_c, we skip the home i. If the correlation between home i and any selected homes is larger than d_c, then we add home i as one of the selected homes. We continue this process until all homes are either skipped or selected one of the homes.

C. Energy Consumption Forecast with Limited Data

In previous design, we consider that every home in smart grids are deployed with smart meters. However, it may not be true in reality. Thus, we also propose an algorithm to improve energy consumption forecast in smart grids when only partial of homes in smart grids are deployed with smart meters. The key idea is to first conduct the energy matching with existing individual homes' energy consumption data we collect in the central server. Then based on the matching results, we know the information of how many homes of different power consumption patterns exist in smart grid. Based on the energy consumption prediction in individual homes with smart meters and aggregated energy consumption in smart grids, we utilize the matching results to predict the energy consumption of smart grids in future.

The high-level idea of our algorithm is that it searches for a shapelet which can separate and remove a subset of time series from the rest of the dataset, then iteratively repeats this search among the remaining data until no data remains to be separated.

As discussed before, an ideal shapelet has the ability to divide a dataset **D** into two groups of time series, D_A and D_B. D_A consists of the time series that have subsequences similar to while D_B contains the rest of the time series in **D**. Simply stated, we expect the mean value of $sdist(S, D_A)$ to be much smaller than the mean value of $sdist(S, D_B)$. Since we ultimately use a distance map that contains distance vectors to cluster the dataset, the larger the gap between these two means of these distances vectors, the better. We use the algorithm to extract shapelets. In essence, this algorithm can be seen as a greedy search algorithm which attempts to maximize the separation gap between two subsets of D. This separation measure is formally encoded in the following equation:

$$gap = \mu_B - \sigma_B - (\mu_A + \sigma_A) \tag{13}$$

In Equation 13, μ_A and μ_B represent $mean(sdist(S, D_A))$ and $mean(sdist(S, D_B))$ respectively, while σ_A and σ_B represent $std(sdist(S, D_A))$

Algorithm 4 Energy Consumption Prediction with Limited Data

for t = 1 to T_p **do**
  Calculate gap based on Equation 7 and 13;
end for
for t = $starttime1$ to $endtime1$ **do**
  $d(t) = d(t) + power1$;
end for
for t = $starttime2$ to $endtime2$ **do**
  if $t\%(time + period)$ = 0 **then**
    for i = 1 to $time$ **do**
      $d(t) = d(t) + power2$;
    end for
  end if
end for
if b_c = 1 **then**
  for i = 1 to m **do**
    $d(t) = d(t) + e(t)$;
  end for
end if

and $std(sdist(S, D_B))$, respectively. In our algorithm, we consider all subsequences of the time series as candidate shapelets and compute their distance vectors. We can represent a distance vector as a schematic line. Then we search these lines for the location that maximizes the gap function introduced in Equation 13. We refer to this point as dt. Points to the left of dt represent $sdist(S, D_A)$, while points to the right correspond to $sdist(S, D_B)$

Once we know the gap scores for all the subsequences of a time series, we add the subsequence with maximum gap score in the set of shapelets. Given that we have selected a u-shapelet, we do not want subsequences similar to it to be selected as shapelets in subsequent iterations. Thus we remove the time series that have subsequences similar to the shapelet from the dataset and use only the remaining dataset to search for the next shapelet.

The detailed design of energy consumption prediction is shown in Algorithm 4. For polynomial fitting, we can calculate gap with $\{T_p, a_0, \cdots, a_n\}$ based on Equation 7 and 13. For energy consumption pattern abstraction, we can calculate high energy consumption patterns in a short time and periodical energy consumption over a long time with $\{power2, starttime2, endtime2, period, time\}$. For utilization of consumption correlation, we can use b_c to detect whether we utilize correlation of energy consumption. For the limited data, we apply energy matching algorithm for energy consumption pattern from given individual homes and aggregated energy consumption data. Finally, with the data from four components, we calculate the predicted energy consumption in smart grids.

D. Time Complexity Analysis

In this section, we analyze the complexity of M-Pred at local smart meters in the following three stages.

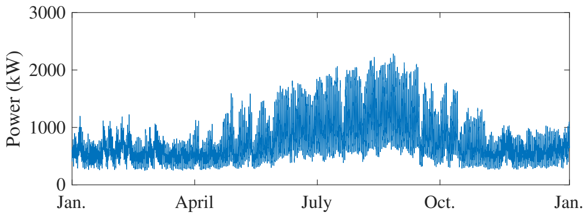

Figure 7: Aggregated power consumption over 12 months

i) *Learning Energy Consumption Patterns*. Basic power consumption curve is sketched in time domain. In learning algorithm, Algorithm 1 is applied to generate sampled power consumption changes and time complexity of Algorithm 1 is $O(N)$.

ii) *Energy Consumption Prediction*. Sampled data of power consumption changes are transferred into energy consumption pattern data. In energy consumption prediction, we need to conduct $\tilde{d}(t)$ and run Algorithm 2. The time complexity is $O(NlogN)$ and time complexity for Algorithm 2 is $O(N)$.

iii) *Energy Prediction for Smart Grids with Limited Data*. The limited data from individual homes is used for energy matching and energy consumption prediction in smart grids. Because the number of homes available is limited, thus the communication overhead between homes and central server is low. Time complexity of matching and energy consumption prediction is $O(NlogN)$.

In total, the time complexity of our design is still $O(NlogN)$, which means our design is simple for smart meters.

V. Implementation and Evaluation

In this section, we evaluate the performance of our proposed design. We deploy eGauge power meters at individual homes to collect the energy consumption data every minute. In our simulation, we use the power consumption traces that we collected from 726 homes for more than one year. To make the figure easy to follow, we only show the aggregated power consumption for 12 months in Figure 7. It can be found that the power consumption for different days varies significantly and is higher in summer while lower in winter.

A. Evaluation Baseline and Metrics

Baselines. To verify the prediction accuracy of our approach, we compare our design with three existing approaches: i) NYISO: New York ISO [3], which is the standard of energy consumption prediction in New York State; ii) CASCE: Southern California Edison ISO [2], which is the standard of energy consumption prediction in South California; and iii) CAISO: California ISO [1], which is the standard of energy consumption prediction in California.

For our design, to verify the prediction accuracy of pattern recognition and home clustering, we also compare to i) our design with only pattern recognition (**PR**), which only utilizes aggregated power consumption in smart grids for prediction; ii) our design with perfect home clustering (**PR+All**), which assumes each home is a cluster. The

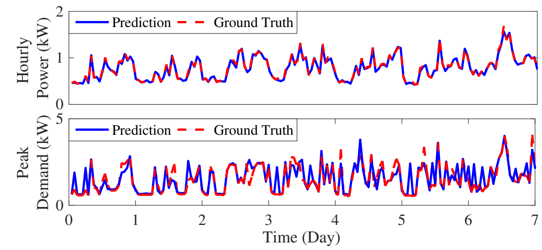

Figure 8: Prediction accuracy of hourly average power and peak demand in a typical home

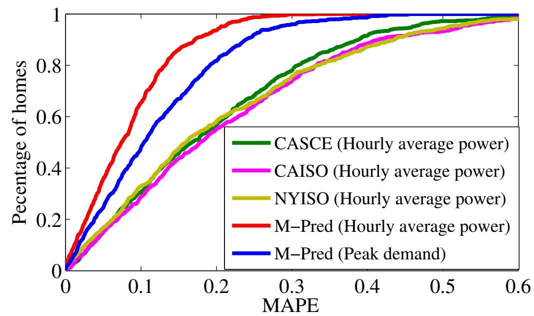

Figure 9: CDF of MAPE for different homes with different methods

prediction accuracy of **PR+All** should be better than our design, however, it is not scalable when the number of homes are huge. In our simulations, we consider **PR+All** as optimal algorithm for prediction accuracy.

Metrics. Because we predict both peak demand and hourly average power consumption in smart grids. Thus, we use two metrics to evaluate the performance of our approach: i) **MAPE of peak demand** and ii) **MAPE of hourly average power**.

B. Energy Consumption Prediction in A Single Home

To enable real-time demand response in individual homes, energy consumption prediction in a single home must be very accurate. In our simulations, we run the pattern recognition algorithm with six months data and conduct the prediction with another six months. The prediction window size in these series of simulations are set as 24 hours.

1) Hourly Average Power Consumption and Peak Demand Prediction: To make the results easy to follow, we only show the prediction results of one typical home for one week in Figure 8. The prediction of hourly average power consumption matches very well with the ground truth. The prediction of peak demand is also very accurate for the most of time. However, the prediction of peak demand is less accurate compared to hourly average power consumption. This is because hourly average power consumption in a single is typical more stable; in the meantime, peak demand is highly dependent on the accurate time of energy consumption events. Thus, it is much more difficult to predict peak demand accurately.

2) Prediction Results for Different Homes: The prediction results of different homes compared to existing approaches are shown in Figure 9. Because existing approaches can only predict hourly average power consumption, we

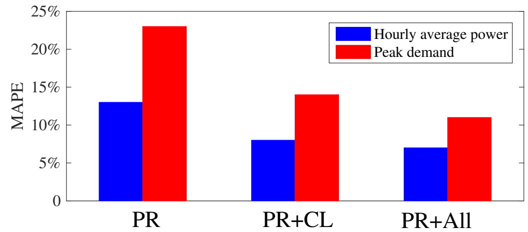

Figure 10: MAPE of aggregated hourly average power and peak demand prediction in smart grids

only show the prediction results of peak demand with our design. For CASCE, CAISO and NYISO, the prediction accuracy is similar and varies from different homes, which is consistent with the results from [7]. The prediction results of hourly average power with our design is much better than existing approaches, 95% of the homes can be predicted with MAPE less than 0.2 and average MAPE for all the homes is 0.08. For the prediction of peak demand in different homes, the accuracy is also very good. 83% of the homes can be predicted with MAPE less than 0.2 and average MAPE for all the homes is 0.11.

C. Energy Consumption Prediction in Smart Grids

Because smart grids need accurate aggregated power consumption prediction, thus we provide the prediction results of aggregated peak demand and hourly power consumption in this section.

1) Prediction with Different Methods: To verify the detailed performance of our design, we compare prediction results with **PR** and **PR+All**. For **PR**, we applied our pattern recognition algorithm with aggregated power consumption in smart grids. No minute-level energy data from individual homes are used for **PR**. For **PR+All**, we assume all the minute-level energy data from individual homes are available in the central server. In reality, it is not practical because large amount of energy data needs to be transmitted from smart meters to central server especially when the number of homes in a smart grid is huge. In this section, we consider **PR+All** as the optimal prediction accuracy we can achieve by utilizing minute-level energy data from individual homes. The MAPE of different methods is shown in Figure 10. We can find that for all three methods, the prediction accuracy of hourly average power is always better than prediction of peak demand, which is similar to prediction accuracy in a single home. Compared to **PR**, the prediction accuracy of hourly average power consumption and peak demand with M-Pred is both around 40% better and very close to optimal results of **PR+All**. Therefore, our design is well balanced between prediction accuracy and communication overhead.

2) Prediction with Limited Available Homes: Considering not every home in smart grids is deployed with smart meters, we evaluate the performance of our design with minute-level energy data from different number of homes. The results are shown in Figure 11. The X-axis is the percentage of homes that are deployed with smart meters in a smart grid; and Y-axis is the MAPE value of prediction results. We can find

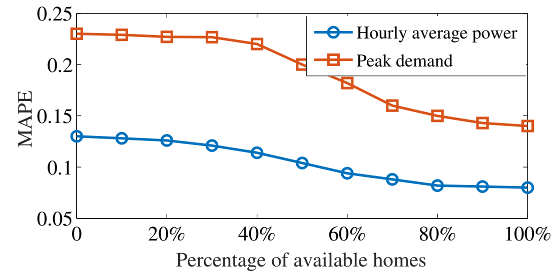

Figure 11: MAPE of aggregated hourly average power and peak demand prediction with different percentages of homes with minute-level energy data

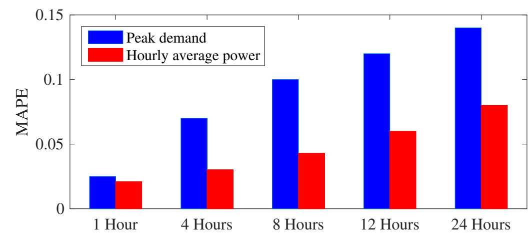

Figure 12: MAPE of aggregated hourly average power and peak demand prediction with different window size

that the prediction accuracy of both hourly average power consumption and peak demand increase with minute-level energy data from more homes. For prediction of hourly average power consumption, the accuracy increases more slowly with percentage of available homes. This is because in residential homes, energy consumption is mostly high in the morning, low in the day time and high in the evening. Therefore, the curve of hourly average power consumption over time is similar for different homes. At the mean time, prediction accuracy of peak demand only increases significantly when the percentage of available homes reaches 40%. This is because peak demand in different homes are usually more dependent on homeowners' behavior patterns, thus, only with enough energy data from individual homes, the prediction accuracy of peak demand will be improved.

D. Impact of Window Size

In this section, we investigate the impact of window size on the prediction accuracy of hourly average power consumption and peak demand. We apply our design with window size of 1 hour, 4 hours, 8 hours, 12 hours and 24 hours. The results are shown in Figure 12. With larger window size, the consumption events of homeowners' are more unpredictable. Thus, the prediction accuracy of both hourly average power consumption and peak demand decreases with larger window size. For 1 hour window size, the prediction accuracy of our design is extremely high with MAPE around 0.02. Therefore, our design can provide very accurate power consumption prediction for real-time demand response in smart grids. Similar to results in previous sections, prediction accuracy of peak demand decrease faster with large window size because the peak demand is more unpredictable.

VI. Related Work

Our work is related to two areas of previous work: demand forecast and peak demand:

Demand Forecast. Research on electricity demand forecast includes long-term and medium-term prediction for utility planning and maintenance purposes, and short-term forecast for economic scheduling [4, 13]. In this paper, we focus on the short-term demand forecast. Related work on demand forecast includes three types of methods: simple averaging models [1, 2, 3]; statistical models (e.g., regression [15] and time series [6, 10]); and machine learning techniques(e.g., Artificial Neural Networks (ANNs) [11, 17] and pattern matching [16, 18]). However, existing forecast techniques only conduct forecast with aggregated power consumption in samrt grids. In this paper, we show that with detailed power consumption in individual homes collected from smart meters, power consumption patterns in each home can significantly help the demand forecast in smart grids.

Peak Demand. There are many works on modifying the elastic load components of common household appliances to reduce peak demand [12]. In [19], a novel demand response mechanism is proposed to exploits appliance elasticity to decrease peak loads. A real-time distributed deferrable load control algorithm is proposed to reduce the peak load by shifting the power consumption of deferrable loads to periods with high renewable generation [8]. To support different approaches on flattening peak demand in smart grids, we present peak demand forecast in this paper for the first time. The simulation results show that our design can significantly improve prediction accuracy of peak demand in smart grids.

VII. Conclusion

To the best of our knowledge, this is the first work to utilize the detailed power consumption in individual homes to help power consumption prediction in smart grids. We show that the detailed power consumption patterns in each home can significantly improve prediction accuracy of power consumption in smart grids. In this paper, we propose M-Pred to learn energy consumption pattern of individual homes from their energy consumption data and then utilize these patterns to predict the power consumption in smart grids. Our design consists of three parts: i) energy consumption patterns recognition in a single home; ii) energy consumption prediction in smart grids; iii) energy consumption prediction with limited data from individual homes. Finally, we analyze the performance and complexity of M-Pred. We conducted extensive system evaluations with 726 homes' minute-level power consumption data for more than 1 year. The results show that our design can provide accurate real-time energy consumption with negligible errors (e.g., Mean Absolute Percentage Error is 2.12%).

VIII. Acknowledgement

This project is supported by NSF grant CNS-1503590 and UMBC COEIT Strategic Plan Implementation Grant.

References

[1] California ISO. CAISO demand response resource user guide. *Tech. Rep.*, 2007.

[2] Southern California Edison. 10 day average baseline and "day-off" adjustment. *Tech. Rep.*, 2011.

[3] New York ISO. Emergency demand response program manual. *Tech. Rep.*, 2013.

[4] H. K. Alfares and M. Nazeeruddin. Electric load forecasting: literature survey and classification of methods. *International Journal of Systems Science*, 33(1):23–34, 2002.

[5] S. Aman, Y. Simmhan, and V. K. Prasanna. Holistic measures for evaluating prediction models in smart grids. *IEEE Transactions on Knowledge and Data Engineering*, 27(2):475–488, 2015.

[6] S.-C. Chan, K. M. Tsui, H. Wu, Y. Hou, Y.-C. Wu, and F. F. Wu. Load/price forecasting and managing demand response for smart grids: Methodologies and challenges. *IEEE Signal Processing Magazine*, 29(5):68–85, 2012.

[7] M. Frincu, C. Chelmis, M. U. Noor, and V. Prasanna. Accurate and efficient selection of the best consumption prediction method in smart grids. In *IEEE International Conference on Big Data*, 2014.

[8] L. Gan, A. Wierman, U. Topcu, N. Chen, and S. H. Low. Real-time deferrable load control: Handling the uncertainties of renewable generation. In *e-Energy*, 2013.

[9] S. Haben, J. Ward, D. V. Greetham, C. Singleton, and P. Grindrod. A new error measure for forecasts of household-level, high resolution electrical energy consumption. *International Journal of Forecasting*, 30(2):246 – 256, 2014.

[10] M. T. Hagan and S. M. Behr. The time series approach to short term load forecasting. *IEEE Transactions on Power Systems*, 2(3):785–791, 1987.

[11] H. S. Hippert, C. E. Pedreira, and R. C. Souza. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems*, 16(1):44–55, 2001.

[12] H. Hlavacs, K. Hummel, R. Weidlich, A. Houyou, A. Berl, and H. Meer. Energy efficiency in future home environments: A distributed approach. *Home Networking*, 256:69–84, 2008.

[13] Z. Huang, H. Luo, D. Skoda, T. Zhu, and Y. Gu. E-sketch: Gathering large-scale energy consumption data based on consumption patterns. In *IEEE International Conference on Big Data*, 2014.

[14] Z. Huang, T. Zhu, H. Lu, and W. Gao. Accurate power quality monitoring in microgrids. In *IPSN*, 2016.

[15] J. Z. Kolter and J. Ferreira Jr. A large-scale study on predicting and contextualizing building energy usage. 2011.

[16] F. Martinez Alvarez, A. Troncoso, J. C. Riquelme, and J. S. Aguilar Ruiz. Energy time series forecasting based on pattern sequence similarity. *IEEE Transactions on Knowledge and Data Engineering*, 23(8):1230–1243, 2011.

[17] S. D. Ramchurn, P. Vytelingum, A. Rogers, and N. R. Jennings. Putting the "smarts" into the smart grid: a grand challenge for artificial intelligence. *Communications of the ACM*, 55(4):86–97, 2012.

[18] W. Shen, V. Babushkin, Z. Aung, and W. L. Woon. An ensemble model for day-ahead electricity demand time series forecasting. In *e-Energy*, 2013.

[19] P. Srikantha, S. Keshav, and C. Rosenberg. Distributed control for reducing carbon footprint in the residential sector. In *SmartGridComm*, 2012.

[20] M. Weiss, A. Helfenstein, F. Mattern, and T. Staake. Leveraging smart meter data to recognize home appliances. In *PerCom*, 2012.