Signal Disaggregation via Sparse Coding with Featured Discriminative Dictionary

Bingsheng Wang[1], Feng Chen[1], Haili Dong[1], Arnold P. Boedihardjo[2], Chang-Tien Lu[1]

[1]Virginia Tech, Falls Church, VA 22043
[2]U. S. Army Corps of Engineers, Alexandria, VA 22315
claren89, chenf, hailid@vt.edu, arnold.p.boedihardjo@usace.army.mil, ctlu@vt.edu

Abstract—**As the issue of freshwater shortage is increasing daily, it's critical to take effective measures for water conservation. Based on previous studies, device level consumption could lead to significant conservation of freshwater. However, current smart meter deployments only produce low sample rate aggregated data. In this paper, we examine the task of separating whole-home water consumption into its component appliances. A key challenge is to address the unique features of low sample rate data. To this end, we propose Sparse Coding with Featured Discriminative Dictionary (SCFDD) by incorporating inherent shape and activation features to capture the discriminative characteristics of devices. In addition, extensive experiments were performed to validate the effectiveness of SCFDD.**

I. INTRODUCTION

The scarcity of potable water is one of the most critical challenges facing the world. The statistics shown in Nature 2010 [1] states that about 80% of the world's population lives in short of potable water. Furthermore, according to the California Department of Water Resources, without more water supplies by 2020, the region will suffer a deficiency nearly as much as the total amount consumed today [2]. At the global level, the existing freshwater is only enough to extend out as much as 60 or 70 years [3]. Urban water consumption contributes to 50%-80% of public water supply systems and 26% of whole usage in the US[4].

Studies have shown that device-level water usage information is crucial for establishing effective conservation strategies [5-7]. This paper specifically considers the task of disaggregating residential water consumption targeting for conservation. Consumption disaggregation refers to the process of separating aggregated smart meter readings into the consumption of its devices, such as toilet, shower and washer.

Recently, water consumption disaggregation has become an important topic to explore solutions for water conservation. Most previous studies focus on sensing the open/close pressure waves of devices to identify the signatures for separation. These methods are capable of achieving more than 90% in accuracy measure. However, they depend on high sample rate (typically 1 kHz) data sources to analyze individual device features. The widely deployed smart meters only produce low sample rate (as low as 1/900 Hz) readings to ensure reliable data transmission. It's critical to design an effective algorithm to disaggregate low sample rate aggregated data.

The challenge is to cope with the issues caused by low sample rate data sources, since it's impossible to identify the open/close signatures of devices from the data at such a low resolution [8, 9]. This inspires us to model new "signatures" for devices from low sample rate data. Through integrating the water usage cycles (or duration) with interval based consumption trend (per interval is 15 min), we propose to use shape features as one "signature" for distinguishing devices. In addition, another "signature" is defined as the amplitude of activations, indicating the consumption of device. For instance, with respect to consumption, the "signature" of toilet is 1~5 gallons while that of shower is 6~30 gallons. We then integrate the inherent shape and activation features into Sparse Coding with Featured Discriminative Dictionary (SCFDD) to separate the whole-home data into its components. The contributions of this paper are as follows:

- **In-depth study on the inherent shape features of aggregated data:** We inspect the shape features for every device, and propose an efficient algorithm to extract the features based on First-order Relation.
- **Comprehensive examination of the amplitude of activations:** The consumption characteristics of devices are investigated and considered as the discriminative features, which are depicted by the amplitude of activations.
- **Extension of Sparse Coding with Featured Discriminative Dictionary (SCFDD):** SCFDD is proposed by incorporating the inherent shape and activation features into the sparse coding model, which can effectively perform water consumption disaggregation task.
- **Extensive experiments for illustrating the effectiveness of SCFDD:** We demonstrated the effectiveness of SCFDD by comparing it with other models in both whole-home and device level measures.

This paper is organized as follows. Section II covers the background and related work. The sparse coding with featured discriminative dictionary model is introduced in Section III. The effectiveness of the proposed method is validated with extensive experiments in Section IV. Our work is summarized in Section V.

II. BACKGROUND & RELATED WORK

We provide the notations and concepts, and introduce sparse coding based approaches.

IEEE computer society

A. Notations and Concepts

Assume we are given d different devices, such as toilet, shower, and washer. For each device $i = 1, \ldots, d$, we have a water consumption matrix $\mathbf{Y}_i \in \mathbb{R}^{T \times m}$, where T is the number of intervals in one day, and m is the number of days. The jth column of $\mathbf{Y}_i$, denoted by $\mathbf{y}_{ij}$, is the jth day's water usage for device i, where $j = 1, \ldots, m$. The ℓth element of $\mathbf{y}_{ij}$, denoted by $y_{ij\ell}$, $\ell = 1, \ldots T$, is the ℓth interval's water usage in the jth day for device i. We denote aggregated water consumption over all devices as $\overline{\mathbf{Y}} \equiv \sum_i^d \mathbf{Y}_i$. Thus, each column of $\overline{\mathbf{Y}}$ indicates one day's aggregated water consumption of all devices. In the training process, we have individual device consumption data, $\mathbf{Y}_1, \ldots, \mathbf{Y}_d$. At test time, we only have a new set of aggregated consumption data, $\overline{\mathbf{Y}}'$, and the goal is to disaggregate it into its components, $\mathbf{Y}_1', \ldots, \mathbf{Y}_d'$.

Sparse coding for source separation [10] and energy disaggregation [11] is to learn device based dictionary, and then separate aggregated signals into individual devices. Formally, for device i, the data matrix is denoted as $\mathbf{Y}_i$, $\mathbf{Y}_i = \mathbf{B}_i \mathbf{A}_i + \boldsymbol{\epsilon}_i$, where $\mathbf{B}_i \in \mathbb{R}^{T \times n_i}$ is the *dictionary* for device i, and the columns of $\mathbf{B}_i$ contain a set of n_i basis functions; $\mathbf{A}_i \in \mathbb{R}^{n_i \times m}$ is named as the *activations* of the device i's dictionary; $\boldsymbol{\epsilon}_i$ is Gaussian, white noise [12]. The notion of sparseness is achieved by shaping the probability distribution of each element within $\mathbf{A}_i$ to have a high probability at zero.

B. Discriminative Sparse Coding

We discuss the basic idea of discriminative sparse coding in energy disaggregation [11], where the prior on activations of non-negative sparse coding [13] is Laplace distribution. The complete-data likelihood could be converted into the following optimization function [12],

$$\min_{\mathbf{A}_i \geq 0, \mathbf{B}_i \geq 0} \frac{1}{2} \|\mathbf{Y}_i - \mathbf{B}_i \mathbf{A}_i\|_F^2 + \lambda \sum_{jt} a_{ijt}$$
$$\text{subject to } \|\mathbf{b}_{ij}\|_2 \leq 1, j = 1, \ldots, n_i \qquad \text{Eq. (1)}$$

where $\mathbf{Y}_i, \mathbf{A}_i$ and $\mathbf{B}_i$ are defined as in section II.A., $\lambda \in \mathbb{R}_+$ is a regularization parameter. $\|\mathbf{X}\|_F$ is the Frobenius norm, and $\|\mathbf{x}\|_2$ is the ℓ_2 norm. The device based dictionary and activation is learned based on Eq. (1) through executing convex optimization on each variable when holding the other fixed.

With the assumption that the activations estimated from Eq. (1) are the best possible values, and augmented regularized disaggregation error objective function is defined as shown in Eq. (2) to drive dictionary to be discriminative.

$$\tilde{E}_{reg}(\mathbf{Y}_{1:d}, \mathbf{B}_{1:d}, \widetilde{\mathbf{B}}_{1:d}) \equiv \sum_{i=1}^d F(\mathbf{Y}_i, \mathbf{B}_i, \mathbf{A}_i)$$

subject to $\widehat{\mathbf{A}}_{1:d} = \arg\min_{\mathbf{A}_{1:d} \geq 0} F'$, where,

$$F(\mathbf{Y}_i, \mathbf{B}_i, \mathbf{A}_i) = \frac{1}{2} \|\mathbf{Y}_i - \mathbf{B}_i \widehat{\mathbf{A}}_i\|_F^2 + \lambda \sum_{jt} \hat{a}_{ijt}, \qquad \text{Eq. (2)}$$

$$F' = \left\| \overline{\mathbf{Y}} - [\mathbf{B}_1 \quad \cdots \quad \mathbf{B}_d] \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_k \end{bmatrix} \right\|_F^2 +$$

$$\lambda \sum_{ijt} (\mathbf{A}_i)_{ijt}$$

$\mathbf{X}_{1:d}$ is the short hand of $\mathbf{X}_1, \ldots, \mathbf{X}_d$. As the $\mathbf{B}_{1:d}$ bases are the same as learned from Eq. (1), while the $\widetilde{\mathbf{B}}_{1:d}$ bases are discriminatively optimized in order to move $\widehat{\mathbf{A}}_{1:d}$ close to the activations learned from Eq. (1). Finally, the optimization process of Eq. (2) is completed with structured perceptron algorithm [14].

C. Related Work

The primary attempts for disaggregation tasks focus on **power consumption disaggregation**. Nonintrusive Appliance Load Monitoring (NALM) was proposed by Hart [15] for breaking apart residential power consumption, and he indicated that the distinct power consumption signatures of different electrical appliances could be depicted with Finite State Machines (FSM). Recently, Kolter *et al.* proposed Discriminative Disaggregation Sparse Coding (DDSC) for separating low sample rate aggregated power readings [11]. However, the proposed method requires a large amount of training data for training purpose. An unsupervised framework for the disaggregation of low frequency power measurements [16] was proposed by integrating some additional features (e.g. time of day) into Factorial Hidden Markov Model (FHMM).

With respect to **water consumption disaggregation**, Froehlich *et al.* designed pressure-based sensors for water fixtures to identify the activity at individual water fixtures within a home [17]. But this method depends on high sample rate data. A HMM based approach was proposed in [18] to analyze human activities with low sample rate aggregated smart meter readings. However, this method analyses activities in one sequenced data, limiting its ability to parse parallel activities.

There is a lack of models designed for disaggregating low sample rate water consumption. Existing methods are mainly proposed for power consumption disaggregation, which are incapable of analyzing the features of aggregated smart meter readings collected at a low resolution. HMM based method analyses the activities with interval based consumption; however, it has limited ability to estimate the consumption for devices.

III. SPARSE CODING WITH FEATURED DISCRIMINATIVE DICTIONARY

We begin by examining the shape features for low sample rate water consumption. Subsequently, we insepct activation features to characterize the consumption feature for devices. The signal disaggregation is finally performed with Maximum a Posteriori (MAP) estimation.

A. Inherent Shape Features Discovery

Based on domain knowledge, it is suggested that the time duration and consumption trend are distinct across devices [19]. For example, time duration of toilet is around 1~3 minutes, while time duration of shower is around 3~20 minutes. Under the context of this paper, the sample

rate of smart meters is 1/900 HZ, indicating that per reading is 15 min, and there are 96 intervals in one day. We use span and consumption trend to depict shape features.

Definition 1. (Span) For any device i, the span of i, denoted by S_i, is defined as the enumeration of all possible time duration of device i measured as the number of intervals.

Take toilet as an example, there are two possible time duration, 1 or 2. So by Definition 1, $S = \{1,2\}$. Since water consumption is continuous, it's complex to define trend features. For instance, only with respect to 2 intervals, toilet might contain infinite combinations as long as the sum is in a certain range (1~5 gallons [19]), such as

$$\left\{\begin{bmatrix}0.7\\0.8\end{bmatrix},\begin{bmatrix}1.0\\0.5\end{bmatrix},\begin{bmatrix}0.5\\1.0\end{bmatrix},\begin{bmatrix}0.8\\2.5\end{bmatrix},\begin{bmatrix}3.2\\1.7\end{bmatrix},\begin{bmatrix}0.8\\0.8\end{bmatrix},\begin{bmatrix}1.0\\1.0\end{bmatrix},\dots\right\}, \quad \text{Eq. (3)}$$

where each vector shows one possible toilet consumption distributed over 2 intervals (for simplification, we only show the 2 of 96 intervals in one day). Actually, the infinite possibilities cause the problem to be intractable. Thus, it's critical to invent a new method for approximation. Inspecting data shown in Eq. (3), we could intuitively reach the pattern that there exists numerical relation (larger than, equal, or less than) between values in these two intervals. This indicates that the maximum number of consumption trend over z intervals is $3^{z(z-1)/2}$, where z is an integer larger than 0. It grows exponentially with the number of intervals. To simplify this problem, we propose a novel method to approximate the trend of consumption through only considering *first-order relations* (see Definition 2).

Definition 2. (First-order Relation) Given any time series consisting $\mathcal{T}$ real values $v_1, \dots, v_{\mathcal{T}}$, we first use $v_{\min}$ and v_{max} to respectively denote the minimum and maximum values of these $\mathcal{T}$ real values. The first order relation on this time series is defined as,

$$\mathcal{F}(1) = \begin{cases} 1, & if\ \mathcal{T} = 1 \\ 1, & if\ v_1 > v_2, \mathcal{T} > 1 \\ 1, & if\ v_1 = v_2 = v_{max}, \mathcal{T} > 1 \\ 0, & otherwise \end{cases} \quad \text{Eq. (4)}$$

$$\mathcal{F}(t) = \begin{cases} 1, & if\ v_t > v_{t-1}, \mathcal{T} > 1 \\ 0, & if\ v_t < v_{t-1}, \mathcal{T} > 1 \\ \mathcal{F}(t-1), & if\ v_t = v_{t-1}, \mathcal{T} > 1 \end{cases}$$

where $t = 2, \dots, \mathcal{T}$.

For example, all data in Eq. (3) are approximated as,

$$\left\{\begin{bmatrix}0\\1\end{bmatrix},\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}1\\1\end{bmatrix}\right\} \quad \text{Eq. (5)}$$

where the first vector in Eq. (5) indicates that the first element is less than the second element; the second vector indicates the first element is larger than the second element; and the last vector indicates these two elements are the same.

Based on Eq.(4), we can convert all vectors in Eq. (3) to Eq. (5). For a specific device i, we have significantly reduced the complexity from infinity (∞) to $2^{S_{ij}} - 1$, where S_{ij} is the jth element in S_i. Practically, considering general water consumption devices, such as toilet, shower, and washer, the maximum interval based duration is 7 (1 hour and 45 minutes). This indicates that the most complex ordering combination is $2^7 - 1 = 127$, which is certainly acceptable.

Definition 3. (Consumption Mapping) Given any device i and its corresponding span S_i. For $\forall y_{ij} \in Y_i$, $j = 1, \dots, m$, consider any possible time series combination c_{ij}^t of *non-zero* values in y_{ij} while holding original time sequence, where $t \in S_i$ specifies the span of current combination. Then consumption mapping is the process to apply *first-order relation* on c_{ij}^t to get $\mathcal{F}_{c_{ij}^t}$.

If there is water consumption, it is shown as the *non-zero* values in y_{ij}. Thus, in Definition 3, we only consider y_{ij}'s *non-zero* values. Here we use $\mathcal{M}_i$ to denote the consumption mapping results of Y_i.

Definition 4. (Shape Features) Given $\mathcal{M}_i$, the shape features of device i denoted by S_i is the set of unique elements in $\mathcal{M}_i$, $i = 1, \dots, d$.

The shape features of Eq. (3) are shown in Eq. (5). However, each element in the consumption mapping results is associated with one element in the time series set. For instance, the consumption mapping results of Eq. (3) is:

$$\left\{\begin{bmatrix}0\\1\end{bmatrix},\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix},\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}1\\1\end{bmatrix},\dots\right\} \quad \text{Eq. (6)}$$

The proposed *shape features* effectively capture time duration and consumption trend. Based on prior knowledge [19], we can define the span for general devices. Taking shower as an example, time duration 3~20 minutes is able to capture more than 95% situations. Thus, $S_i = \{1,2\}$.

The key to sparse coding is to learn discriminative dictionary for separating whole-home level consumption into appliance level usage. Based on the fact that shape features are the abstract discriminative characteristics of appliances, we incorporate shape features into the construction of dictionary $\mathbf{B}_i^S$ for device i in Algorithm 1.

Algorithm 1 Learn discriminative dictionary based on shape features

input: data matrix for individual device $\mathbf{Y}_i \in \mathbb{R}^{T \times m}$, predefined span for each device S_i, $i = 1, \dots, d$.
output: shape features set and dictionary of particular device $\boldsymbol{\mathcal{S}}_i, \mathbf{B}_i^{\mathcal{S}}$, $i = 1, \dots, d$.
variable: mapping results of every device $\boldsymbol{\mathcal{M}}_i$, the initial values of $\boldsymbol{\mathcal{M}}_i = [\ ], i = 1, \dots, d$.

1   **for** $i = 1$ **to** d
2   **for** $j = 1$ **to** m
3       $\mathcal{Y}_+ \leftarrow \mathbf{y}_{ij} > 0$.
4       $c_{ij}^t \leftarrow$ all possible time series combinations of $\mathcal{Y}_+$ subject to original time sequence, where $t \in S_i$ indicates the span of current combination.
5       $\mathcal{F}_{c_{ij}^t} \leftarrow$ **consumption mapping results** of c_{ij}^t.
6       $\boldsymbol{\mathcal{M}}_i \leftarrow \left[\boldsymbol{\mathcal{M}}_i, \mathcal{F}_{c_{ij}^t} \right]$.
7   ***Extend*** c_{ij}^t as a basis function through filling all other intervals to be zero, and add this basis function at the end of $\mathbf{B}_i^{\mathcal{S}}$.
8   **end for**
9     $\boldsymbol{\mathcal{S}}_i \leftarrow unique(\boldsymbol{\mathcal{M}}_i)$
10    $\mathbf{B}_i^{\mathcal{S}} \leftarrow \mathbf{B}_i^{\mathcal{S}} ./\left\| \mathbf{B}_i^{\mathcal{S}} \right\|_2$.
11 **end for**

In Algorithm 1, function $\boldsymbol{unique(\mathbf{X})}$ is the operation to find element based different vectors from matrix $\mathbf{X}$. The "./" in line 10 means the normalization operation over all basis functions in $\mathbf{B}_i^{\mathcal{S}}$.

As the device based dictionary $\mathbf{B}_i^{\mathcal{S}}$ is extracted from training data $\mathbf{Y}_i$ constrained by shape features $\boldsymbol{\mathcal{S}}_i$, this indicates that $\mathbf{B}_i^{\mathcal{S}}$ is customized for device i, which embodies device based distinctive characters: time duration and consumption trend. Thus, the discovered dictionary $\mathbf{B}^{\mathcal{S}} = \left[\mathbf{B}_1^{\mathcal{S}} \cdots \mathbf{B}_{n_i}^{\mathcal{S}} \right]$ is capable of discriminatively separating whole-home water consumption into its component appliances.

B. Learn Inherent Activation Features and Refine $\boldsymbol{B}^{\mathcal{S}}$

We examine another important discriminative feature relating to the consumption of devices. For example, average water usage of one toilet is about 2.0 gallons, while that of one shower is about 17.8 gallons [19].

Based on the fact that water consumption of individual device basically remain the same value over time [20], the non-zero activations will also remain the same. Because it's activation that determines the consumption given fixed dictionary. In SCFDD, the sparse structure is ensured by l_1 norm.

In addition to zero (or near zero) activations, we are most concerned about the non-zero activations to uncover individual device's consumption. As the learned basis $\mathbf{B}^{\mathcal{S}}$ has already been normalized, we approximate the corresponding activation as the consumption over basis. To determine the probability distribution function accurately captures the non-zero activations, we plot the histogram of activations in Figure 1.
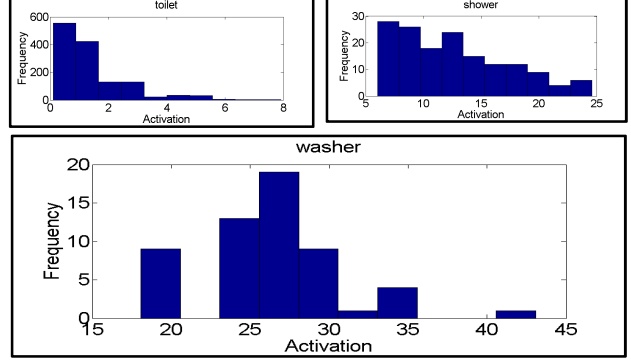

Figure 1. Histogram of non-zero activations.

We found that gamma distribution is suitable to fit most activation features. Since gamma distribution has two parameters (α, β), it has more freedom to scale for various shapes. Thus,

$$P\left(a_{ij\ell} \mid \alpha_i, \beta_i \right) = \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \left(a_{ijl} \right)^{\alpha_i - 1} e^{-\beta_i a_{ijl}} \tag{7}$$

Algorithm 2 Learn activation features and refine dictionary

---

input: dictionary of particular device $\mathbf{B}_i^{\mathcal{S}}$, $i = 1, \dots, d$.
output: activation features α_i, β_i, refined dictionary of each device $\widehat{\mathbf{B}}_i$, $i = 1, \dots, d$.
// **Phase1: Learn activation features**
1   **for** $i = 1$ **to** d
2     $\mathbf{A}_i \leftarrow \operatorname{argmin}_{\mathbf{A}_i \geq 0} \left\| \mathbf{Y}_i - \mathbf{B}_i^{\mathcal{S}} \mathbf{A}_i \right\|_F^2 + \lambda \sum_{j\ell} (\mathbf{A}_i)_{j\ell}$
3     $\alpha_i, \beta_i \leftarrow \operatorname{argmax}_{\alpha_i, \beta_i} P(\alpha_i, \beta_i \mid \mathbf{A}_i)$
4   **end for**
// **Phase2: Refine $\mathbf{B}_i^{\mathcal{S}}$**
5   **for** $i = 1$ **to** d
6   **repeat**
7       $\widehat{\mathbf{B}}_i \leftarrow \operatorname{argmin}_{\mathbf{B}_i \geq 0} \| \mathbf{Y}_i - \mathbf{B}_i \mathbf{A}_i \|_F^2$
8       $\widehat{\mathbf{A}}_i \leftarrow \operatorname{argmin}_{\mathbf{A}_i \geq 0} \left\| \mathbf{Y}_i - \mathbf{B}_i^{\mathcal{S}} \mathbf{A}_i \right\|_F^2 + \lambda \sum_{j\ell} (\mathbf{A}_i)_{j\ell}$
9       $\boldsymbol{\mathcal{O}}_i \leftarrow$ Detect outliers from $\widehat{\mathbf{A}}_i$
10   ***Check*** the span of all basis functions in $\widehat{\mathbf{B}}_i$, and delete those that are out of span of device i.
11      $\widehat{\mathbf{B}}_i \leftarrow \widehat{\mathbf{B}}_i ./\left\| \widehat{\mathbf{B}}_i \right\|_2$.
12   **until** $\boldsymbol{\mathcal{O}}_i = \emptyset$ and convergence.
13 **end for**

To induce the model to prefer the activation fit for the learned gamma distribution, we iteratively remove the abnormal activation values through statistical hypothesis testing.

We detect outliers in gamma distribution through applying the hypothesis testing algorithm proposed by Zerbet and Nikulin [21]. Consequently, we set the spot activations to be zero, and then continue iterating until no abnormal values are detected. The process to learn

activation features and refine dictionary is shown in Algorithm 2.

C. Signal Disaggregation

The generative model for aggregated data,

$$\overline{\mathbf{Y}} = [\mathbf{B}_1 \quad \cdots \quad \mathbf{B}_{n_i}]\begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_{n_i} \end{bmatrix} \tag{8}$$

Given the dictionary $\widehat{\mathbf{B}}$, activation features α_i, β_i, we first estimate $\widehat{\mathbf{A}}$ as

$$\widehat{\mathbf{A}} \leftarrow \operatorname{argmin}_{\mathbf{A} \geq 0}\left\|\overline{\mathbf{Y}} - \overline{\mathbf{B}}\mathbf{A}\right\|_F^2 + \lambda \sum_{ij\ell} \mathbf{A}_{ij\ell} \tag{9}$$

Moreover, for each device i, we detect the outliers from $\widehat{\mathbf{A}}_i$ as $\boldsymbol{\mathcal{O}}_i$. Repeat above process until the union set $\boldsymbol{\mathcal{O}} = \bigcup_i \boldsymbol{\mathcal{O}}_i$ is empty.

In summary, we have presented the model SCFDD for disaggregation task. The inherent shape and activation features are discovered as the differentiate features. The Gamma distribution is used to propel the activations to accommodate the amplitude of consumption.

IV. EXPERIMENT EVALUATION

In this section, we evaluate the performance of the proposed methods based on real-world scenarios.

A. Experiment Setup

1) Datasets: A large-scale dataset containing nearly one million individual water use "events" was collected by Aquacraft from 1,188 residents in 12 study sites (such as Boulder, Colorado and Lompoc, California) [22]. Water consumption for various end uses was measured using compact data loggers and a PC-based flow trace analysis software.

2) Benchmark Methods: We have included the most recent machine learning methodologies for low sample rate data disaggregation: discriminative sparse coding for energy disaggregation and FHMM based methods for energy disaggregation. For the fitness purpose, we compared the proposed methods with five existing related models: discriminative disaggregation sparse coding (DDSC), discriminative disaggregation sparse coding with total consumption priors (DDSC + TCP), discriminative disaggregation sparse coding with group lasso (DDSC + GL), discriminative disaggregation sparse coding with total consumption priors and group lasso (DDSC + TCP + GL), and FHMM. We focused on analyzing three main devices' water consumption: toilet, shower and washer.

3) Evaluation Metrics for Disaggregation Several existing evaluation metrics are applied to evaluate the disaggregation performance. First, we used normalize disaggregation error [23] and accuracy [11] to show the

general disaggregation performance for whole-home level. To evaluate the disaggregation performance for particular device, we used precision, recall and F-measure at device level, where precision is the fraction of disaggregated consumption that is correctly classified while recall is the fraction of true device level consumption that is successfully separated, and F-measure is $2 \times precision \times recall / (precison + recall)$.

B. Performance Evaluation and Comparison

To demonstrate the effectiveness of SCFDD, we compare it with five benchmark methods.

TABLE I shows the disaggregation performance obtained by many prediction methods. SCFDD performed nearly as well or better than all other methods. DDSC and FHMM almost have the same performance. Furthermore, SCFDD was capable of classifying real-world water usage at device level with more than 65% F-measures for toilet, shower, and 30% for washer. Specifically, for shower and washer, the GL extension to DDSC achieved a very poor performance for testing set. This is caused by the fact that GL distorts the model for fitting the testing data.

TABLE I. DISAGGREGATION RESULTS OF METHODS (DDSC: DISCRIMINATIVE DISAGGREGATION SPARSE CODING, FHMM: FACTORIAL HIDDEN MARKOV MODEL, TCP: TOTAL CONSUMPTION PRIOR, GL: GROUP LASSO)

| Method	Device Precision: training set, testing set Recall: training set, testing set F-measure: training set, testing set		
	Toilet	*Shower*	*Washer*
DDSC	52.91%, 31.71%	79.86%, 49.34%	41.62%, 20.73%
	60.46%, 55.87%	66.16%, 22.57%	59.96%, 23.97%
	56.44%, 40.46%	72.37%, 30.97%	49.13%, 22.23%
FHMM	60.72%, 54.39%	69.03%, 63.27%	14.11%, 18.30%
	47.83%, 45.76%	50.45%, 45.99%	42.49%, 44.69%
	53.51%, 49.70%	58.29%, 53.27%	21.19%, 25.97%
DDSC+ TCP	45.27%, 17.68%	59.79%, 39.13%	39.74%, 14.60%
	54.15%, 20.03%	67.47%, 42.16%	53.23%, 16.68%
	49.31%, 18.78%	63.40%, 40.58%	45.51%, 15.57%
DDSC+ GL	40.78%, 27.89%	71.18%, 23.71%	49.46%, 7.19%
	61.80%, 66.99%	42.35%, 4.73%	69.28%, 2.78%
	49.14%, 39.39%	53.10%, 7.88%	57.72%, 4.01%
DDSC+ TCP+ GL	51.45%, 28.44%	82.31%, 24.56%	47.11%, 7.04%
	74.96%, 69.11%	49.12%, 4.48%	70.40%, 2.65%
	61.02%, 40.30%	61.52%, 7.58%	56.45%, 3.85%
SCFDD	84.52%, **58.78%**	93.58%, **74.61%**	58.56%, **39.49%**
	88.06%, **73.81%**	80.03%, **59.72%**	92.89%, **25.43%**
	86.25%, **65.45%**	86.28%, **66.34%**	71.83%, **30.94%**

There is a large gap between training set performance and the testing set performance for washer with respect to precision, recall and F-measure. This is caused by the bias values in the data. Since the frequency of washer is much lower than that of toilet or shower, it's much easier to cause the decrease of performance for washer than the others.

As shown in Figure 2, SCFDD outperformed most other methods with respect to both accuracy and normalized disaggregation error, which indicate the average amount of water usage predicted correctly over

the day. Especially, SCFDD could reach more than 70% in accuracy and less than 0.3 in NDE. As expected, the GL extension degrades both accuracy and NDE of DDSC.
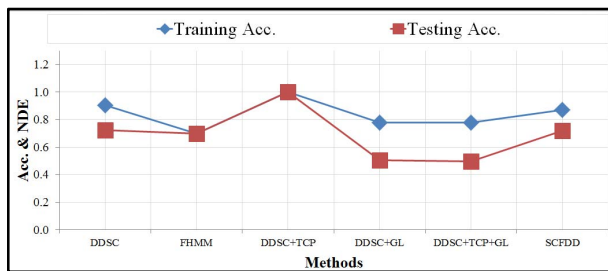
(a). Acc. of the modes.
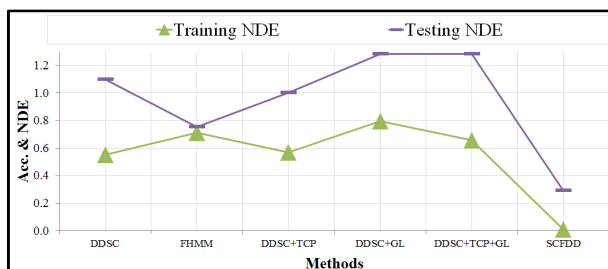
(b). NDE of the modes.

Figure 2. Whole-home performance of models.

DDSC + TCP could achieve a higher accuracy than all others for both training and testing set, due to the consideration of forcing the total estimated amount data to be as close as the true aggregated data. SCFDD could achieve near zero NDE. This is caused by the fact that SCFDD is capable of capturing both shape and amplitude features for correctly estimating the consumption of individual devices.

V.   CONCLUSION

Water consumption disaggregation is a domain where advances in data mining and machine learning could have a significant impact on. In this paper, we provide a concrete solution to the disaggregation of low sample rate smart meter readings. We extract the inherent shape and activation features to discriminatively characterize devices. SCFDD is constructed by incorporating the discovered features into sparse coding for disaggregation task. We apply Gamma distribution to drive the activations of basis to fit the amplitude of consumption. The MAP estimator is employed to predict the consumption of devices. Finally, extensive experiments are performed to demonstrate the effectiveness of SCFDD by comparing it with several benchmark methods.

REFERENCES

[1]   N. Gilbert, "Balancing water supply and wildlife," *Nature,* 2010.
[2]   "California Water Budget," *http://www.water.ca.gov/.*
[3]   P. H. Gleick, "Water Futures: A Review of Global Water Resources Projections1," 1999.

[4]   A. Vickers, *Handbook of Water Use and Conservation: Homes, Landscapes, Industries, Businesses, Farms*, 1st ed.: WaterPlow Press, 2001.
[5]   C. Fischer, "Feedback on household electricity consumption: a tool for saving energy?," *Energy Efficiency,* vol. 1, pp. 79-104, 2008.
[6]   J. Froehlich, L. Findlater, and J. Landay, "The design of eco-feedback technology," in *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, Atlanta, GA, USA, 2010.
[7]   J. E. Froehlich, L. Findlater, M. Ostergren, S. Ramanathan, J. Peterson, I. Wragg, E. Larson, F. Fu, M. Bai, S. N. Patel, and J. A. Landay, "The Design and Evaluation of Prototype Eco-Feedback Displays for Fixture-Level Water Usage Data," in *ACM Annual Conference on Human Factors in Computing Systems*, 2012, pp. 2367-2376.
[8]   E. Larson, J. Froehlich, T. Campbell, C. Haggerty, L. Atlas, J. Fogarty, and S. N. Patel, "Disaggregated water sensing from a single, pressure-based sensor: An extended analysis of HydroSense using staged experiments," *Pervasive and Mobile Computing,* vol. 8, pp. 82-102, 2012.
[9]   J. Froehlich, E. Larson, E. Saba, T. Campbell, L. Atlas, J. Fogarty, and S. Patel, "A Longitudinal Study of Pressure Sensing to Infer Real-World Water Usage Events in the Home," *Pervasive Computing,* pp. 50-69, 2011.
[10]   P. Smaragdis, M. Shashanka, and B. Raj, "A Sparse Non-Parametric Approach for Single Channel Separation of Known Sounds," presented at the Neural Information Processing Systems, 2009.
[11]   J. Z. Kolter, S. Batra, and A. Y. Ng., "Energy Disaggregation via Discriminative Sparse Coding," in *Neural Information Processing Systems*, 2010, pp. 1153-1161.
[12]   B. A. Olshausen and D. J. Fieldt, "Sparse coding with an overcomplete basis set: a strategy employed by V1," *Vision Research,* vol. 37, pp. 3311-3325, 1997.
[13]   P. O. Hoyer, "Non-negative sparse coding," in *IEEE Workshop on Neural Networks for Signal Processing*, 2002, pp. 557-565.
[14]   M. Collins, "Discriminative training methods for hidden Markov models: theory and experiments with perceptron algorithms," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, 2002, pp. 1-8.
[15]   H. G.W., "Nonintrusive appliance load monitoring," *Proceedings of the IEEE,* vol. 80, pp. 1870-1891.
[16]   H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised Disaggregation of Low Frequency Power Measurements," in *SIAM Conference on Data Mining*, 2011, pp. 747-758.
[17]   J. Froehlich, E. Larson, T. Campbell, C. Haggerty, J. Fogarty, and S. N. Patel, "HydroSense: Infrastructure-Mediated Single-Point Sensing of Whole-Home Water Activity," in *Proceedings of the 11th International Conference on Ubiquitous Computing*, 2009, pp. 235-244.
[18]   F. Chen, J. Dai, B. Wang, S. Sahu, M. Naphade, and C.-T. Lu, "Activity Analysis Based on Low Sample Rate Smart Meters," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2011, pp. 240-248.
[19]   P. Roberts, "Yarra Valley Water 2004 Residential End Use Measurement Study," *Melbourne, Yarra Valley Water,* 2004.
[20]   "Water Usage," in *College of Agriculture, University of Kentucky, http://www.ca.uky.edu/enri/pubs/usage.pdf*, 1999.
[21]   M. J. Nooghabia, H. J. Nooghabia, and P. Nasirib, "Detecting Outliers in Gamma Distribution," *Communications in Statistics - Theory and Methods,* vol. 39, pp. 698-706, 2010.
[22]   P. Mayer, W. DeOreo, E. Opitz, J. Kiefer, W. Davis, and B. Dziegielewski, " Residential End Uses of Water," *American Water Works Research Foundation,* 1999.
[23]   J. Z. Kolter and T. Jaakkola, "Approximation Inference in Additive Factorial HMMs with Application to Energy Disaggregation," in *International Conference on Artifical Intelligence and Statistics*, 2012, pp. 1472-1482.