Signal Disaggregation via Sparse Coding with Featured Discriminative Dictionary

Bingsheng Wang¹, Feng Chen¹, Haili Dong¹, Arnold P. Boedihardjo², Chang-Tien Lu¹
¹Virginia Tech, Falls Church, VA 22043
²U. 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.
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 $Y_i \in \mathbb{R}^{T \times m}$, where $T$ is the number of intervals in one day, and $m$ is the number of days. The $j$th column of $Y_i$, denoted by $y_{ij}$, is the $j$th day’s water usage for device $i$. We have a set of $n_i$ basis functions $\mathcal{A}_i^{(n_i \times m)}$ and a sparse coding parameter $\epsilon_i$. We denote aggregated water consumption over all devices as $\tilde{Y} = \sum_{i}^{d} Y_i$. Thus, each column of $\tilde{Y}$ indicates one day’s aggregated water consumption of all devices. In the training process, we have individual device consumption data, $Y_1, \ldots, Y_d$. At test time, we only have a new set of aggregated consumption data, $\tilde{Y}'$, and the goal is to disaggregate it into its components, $Y_1, \ldots, Y_d$.

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 by $Y_i = \mathbf{B}_i \mathbf{A}_i + \mathbf{e}_i$, 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; $\mathbf{e}_i$ 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],

$$
\min_{\mathbf{A}_1 \geq 0, \mathbf{B}_1 \geq 0} \frac{1}{2} \left\| Y_i - \mathbf{B}_1 \mathbf{A}_1 \right\|_F^2 + \lambda \sum_{j=1}^{n_i} a_{ij}\quad \text{Eq. (1)}
$$

subject to $\left\| \mathbf{B}_1 \right\|_2 \leq 1, j = 1, \ldots, n_i$

where $Y_i, \mathbf{A}_i$ and $\mathbf{B}_i$ are defined as in section II.A., $\lambda \in \mathbb{R}_+$ is a regularization parameter, $\left\| \mathbf{X} \right\|_F$ is the Frobenius norm, and $\left\| \mathbf{x} \right\|_2$ is the $\ell_2$ 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,

$$
E_{\text{reg}}(Y_1, \ldots, Y_d, \mathbf{B}_1, \ldots, \mathbf{B}_d) = \sum_{i=1}^{d} F(Y_i, \mathbf{B}_i, \mathbf{A}_i)
$$

subject to $\mathbf{A}_1 \geq 0 = \arg \min_{\mathbf{A}_1 \geq 0} F'$, where,

$$
F(Y_i, \mathbf{B}_i, \mathbf{A}_i) = \frac{1}{2} \left\| Y_i - \mathbf{B}_i \mathbf{A}_i \right\|_F^2 + \lambda \sum_{j=1}^{n_i} a_{ij}, \quad \text{Eq. (2)}
$$

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

$X_{1:d}$ is the short hand of $X_1, \ldots, X_d$. As the $\mathbf{B}_{1:d}$ bases are the same as learned from Eq. (1), while the $\mathbf{B}_{1:d}$ bases are discriminatively optimized in order to move $\mathbf{A}_{1:d}$ 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 inspect 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 approximate the trend of consumption through only one element. Typically, 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.

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 ($\infty$) to $2^{S_i} - 1$, where $S_i$ is the $j$th 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). Thus, it’s critical to invent a new method to approximate the 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

$$\left\{ [0.7, 1.0], [0.5, 0.8], [3.2, 0.8], [1.0, \ldots] \right\}, \quad \text{Eq. (3)}$$

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 $3^{(z-1)/2}$, 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 $T$ real values $v_1, \ldots, v_T$, we first use $v_{\min}$ and $v_{\max}$ to respectively denote the minimum and maximum values of these $T$ real values. The first order relation on this time series is defined as,

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

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

where $t = 2, \ldots, T$.

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

$$\left\{ [0], [1], [0] \right\} \quad \text{Eq. (5)}$$

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.

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:

$$\left\{ [0], [1], [0], [1], [0], [1], [1], \ldots \right\} \quad \text{Eq. (6)}$$

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 $\mathbf{B}_i$ for device $i$ in Algorithm 1.
input: data matrix for individual device $Y_i \in \mathbb{R}^{T \times m}$, predefined span for each device $S_i$, $i = 1, \ldots, d$. 
output: shape features set and dictionary of particular device $B^S_i$, $i = 1, \ldots, d$. 
variable: mapping results of every device $M_i$, the initial values of $M_i = \emptyset$, $i = 1, \ldots, d$. 

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

Algorithm 1: Learn features and refine dictionary

In Algorithm 1, function $\text{unique}(X)$ is the operation to find element based different vectors from matrix $X$. The “/” in line 10 means the normalization operation over all basis functions in $B^S_i$.

As the device based dictionary $B^S_i$ is extracted from training data $Y_i$ constrained by shape features $S_i$, this indicates that $B^S_i$ is customized for device $i$, which embodies device based distinctive characters: time duration and consumption trend. Thus, the discovered dictionary $B^S = [B^S_1 \cdots B^S_d]$ is capable of discriminatively separating whole-home water consumption into its component appliances.

B. Learn Inherent Activation Features and Refine $B^S$

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 $B^S_i$ 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.

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

$$P(a_{ij} | \alpha_i, \beta_i) = \frac{\alpha_i^{a_i}}{\Gamma(\alpha_i)} e^{-\beta_i a_{ij}}$$

(7)

Algorithm 2: Learn activation features and refine dictionary

input: dictionary of particular device $B^S_i$, $i = 1, \ldots, d$. 
output: activation features $\alpha_i$, $\beta_i$, refined dictionary of each device $B_i$, $i = 1, \ldots, d$. 

// Phase1: Learn activation features
1. for $i = 1$ to $d$
2.   $A_i \leftarrow \text{argmin}_{A_i} \|Y_i - B^S_i A_i\|_F^2 + \lambda \sum_{j \in E} (A_i)_{ji}$
3.   $\alpha_i, \beta_i \leftarrow \text{argmax}_{\alpha_i, \beta_i} P(a_{ij} | \alpha_i, \beta_i, A_i)$
4. end for

//Phase2: Refine $B^S_i$
5. for $i = 1$ to $d$
6. repeat
7.   $\tilde{B}_i \leftarrow \text{argmin}_{B_i} \|Y_i - B_i A_i\|_F^2$
8.   $\tilde{A}_i \leftarrow \text{argmin}_{A_i} \|Y_i - B^S_i A_i\|_F^2 + \lambda \sum_{j \in E} (A_i)_{ji}$
9.   $O_i \leftarrow$ Detect outliers from $\tilde{A}_i$
10. $\tilde{B}_i \leftarrow \tilde{B}_i / \|\tilde{B}_i\|_2$.
11. until $O_i = \emptyset$ and convergence.
12. 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...
We have included the most disaggregation error [23] and accuracy [11] to show the disaggregation performance. First, we used normalize existing evaluation metrics are applied to evaluate the devices’ water consumption: toilet, shower and washer. Total consumption priors and group lasso (DDSC + TCP + GL) discriminative models: discriminative disaggregation sparse coding energy disaggregation and FHMM based methods for rate data disaggregation: discriminative sparse coding for software.

Compact data loggers and a PC-based flow trace analysis consumption for various end uses was measured using 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.

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.

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

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 in Denver 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 $2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$.

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.

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.

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 features into sparse coding for disaggregation task. We extract the inherent shape and activation features to discriminatively characterize devices. 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.

Figure 2. Whole-home performance of models.

(a). Acc. of the modes.

(b). NDE of the modes.

Figure 2. Whole-home performance of models.

REFERENCES