Deep Sparse Coding Based Recursive Disaggregation Model for Water Conservation

Haili Dong, Bingsheng Wang, Chang-Tien Lu
Department of Computer Science
Virginia Tech, United States
{hailid, claren89, ctlu}@vt.edu

Abstract

The increasing demands on drinkable water, along with population growth, water-intensive agriculture and economic development, pose critical challenges to water sustainability. New techniques to long-term water conservation that incorporate principles of sustainability are expected. Recent studies have shown that providing customers with usage information of fixtures could help them save a considerable amount of water. Existing disaggregation techniques focus on learning consumption patterns for individual devices. Little attention has been given to the hierarchical decomposition structure of the aggregated consumption. In this paper, a Deep Sparse Coding based Recursive Disaggregation Model (DSCRDM) is proposed for water conservation. We design a recursive decomposition structure to perform the disaggregation task, and introduce *sequential set* to capture its characteristics. An efficient and effective algorithm *deep sparse coding* is developed to automatically learn the disaggregation architecture, along with discriminative and reconstruction dictionaries for each layer. We demonstrated that our proposed approach significantly improved the performance of the benchmark methods on a large scale disaggregation task and illustrated how our model could provide practical feedbacks to customers for water conservation.

1 Introduction

According to the 2012 statistics reported by [WHO, 2012] and [UNICER, 2012], more than 780 million people live in a shortage of portable water and about 2.5 billion people are living in areas where it is difficult or even impossible to access safe sanitation facilities. It's estimated that 2.3 billion people will be living without access to basic water needs for drinking, cooking and sanitation in 2025 [Concern-Worldwide, 2008]. To ensure that the world's water resources can be sustained for future generations, water conservation becomes an increasingly important research topic.

Generally, there are two main approaches to achieve the goal of water conservation: deploying infrastructures for efficient use of water, or reducing water demand by changing consumption habits. The first strategy will inevitably incur a large investment to develop new techniques for intelligent use of water, while the second strategy is more economical and mainly depends on customers' consumption behaviours. Recent studies have shown that detailed feedbacks on consumption patterns would largely affect customers' behaviours and ultimately save consumption up to 20 percent [Fischer, 2008; Froehlich *et al.*, 2010; Houde *et al.*, 2013]. We intend to employ the advances of artificial intelligence to separate aggregated water consumption into component devices, and provide device level usage to users for water conservation.

Water disaggregation is an emerging research topic, which involves taking an aggregated water consumption, for example, the total smart meter readings of a house, and decomposing it into the usages of different water fixtures. Recently, significant research efforts have been attracted to detect the "open" and "close" operations of appliances based on *high sample rate* sensing data. Froehlich *et al.* proposed HydroSense, a pressure based sensor, to identify activities at individual water fixtures within a home and estimate the amount of fixture level water usage [Froehlich *et al.*, 2009]. Furthermore, an extensive study of the HydroSense technology was conducted by Larson *et al.* through a comprehensive analysis of valve- and fixture-level events identification [Larson *et al.*, 2012]. In 2011, Froehlich *et al.* performed a longitudinal study of pressure sensing and concentrated on inferring real-world water usage events in the home [Froehlich *et al.*, 2011]. Srinivasan *et al.* designed flow and motion sensor signatures based clustering techniques to identify unique water fixtures [Srinivasan *et al.*, 2011]. Parson *et al.* used Hidden Markov Model for energy disaggregation in an iterative manner [Parson *et al.*, 2011]. However, all of these studies depend on *high sample rate* (1Hz$\sim$1Khz) sensing data. It is not practical to widely deploy *high sample rate* smart meters in real world due to privacy and reliable data transmission concerns.

This motivates another major stream for water disaggregation research with the usage of *low sample rate* (about 1/900 Hz) sensing data. Kolter *et al.* proposed discriminative sparse coding for low sample rate energy signal disaggregation, and applied structured prediction to refine the basis functions to minimize the regularized disaggregation error [Kolter *et al.*, 2010]. However, the accuracy of this model will decease when common basis functions are shared by many devices.

A FHMM-based unsupervised disaggregation framework is presented in [Kim *et al.*, 2011] to incorporate additional features, such as time of day and dependency between appliances. However, it's impractical to collect such prior knowledge especially when handling data collected from a large scale area. Wang *et al.* incorporated the shape and activation characteristics into sparse coding based dictionary learning, and achieved acceptable performance [Wang *et al.*, 2012]. But only three devices were analysed since it's impractical to manually examine the spans and shapes for all devices, especially when the number of devices is large.

In this paper, we look specifically at the task of *low sample rate* water consumption disaggregation. Since existing approaches are all lack of the mechanisms to learn the decomposition structure, we conduct the first study on the representation learning of the disaggregation process. Through designing a hierarchical recursive structure, we can control the disaggregation sequence. Considering the problems that the learnable parameters of discriminative sparse coding model will significantly increase with the number of devices, several critical issues will emerge, such as over-fitting due to excessive complexity and slow convergence caused by vast local optima. It's of significant importance to perform separation in a structural manner targeting for reducing the size of dictionaries. We disaggregate consumption of each device from total consumption step by step, and learn a much smaller discriminative dictionary to distinguish the current device from all other residual devices in each step. Suppose there is a total consumption of k devices, and there are N bases for each device. If we learn the dictionary for separation using discriminative sparse coding, there will be a dictionary of $k \times N$ for distinguishing the k devices simultaneously. Generally, there are more than 10 identical water appliances in each household, i.e., $k > 10$. On the other hand, using the recursive structure, we only need to learn a dictionary of $2 \times N$ for the separation between the current device and the remainders. An effective algorithm is designed for the discovery of the optimal separation sequence with the objective of minimizing the *recursive disaggregation error* respectively for each layer. In summary, the primary contributions of this paper are as follows:

- **Construction of the hierarchical decomposition structure:** We investigate the advantages of representation learning for disaggregation, and introduce *sequential set* to formalize the recursive structure.
- **Design of a Deep Sparse Coding based Recursive Disaggregation Model (DSCRDM):** DSCRDM is proposed to incorporate the separation architecture into the sparse coding model, and the *recursive disaggregation error* is presented to perform optimization.
- **Efficient and effective learning algorithm for the discovery of separation sequence:** We propose the algorithm *deep sparse coding* by relaxing the exact constraints to learn the layer level discriminative and reconstruction dictionaries sequentially.
- **Extensible experiments to validate the effectiveness of DSCRDM:** We demonstrated that DSCRDM outperformed other models in both whole home and device level measures in a large scale real-word disaggregation task.

2 Preliminaries

In this section, we introduce the notations and discriminative disaggregation sparse coding.

Notations and Concepts Assume there are k water fixtures, such as faucet and irrigation, in a household. For $\forall i = 1, 2, \cdots, k$, there exists a consumption matrix $\mathbf{C}_i \in \mathbb{R}^{T\times M}$, where T is the number of intervals in one day, and M indicates the number of days. The j^{th} column of $\mathbf{C}_i$, denoted as $\mathbf{C}_{ij}$, represents the consumption of device i in the j^{th} day, where $j = 1, 2, \cdots, M$. The l^{th} element of $\mathbf{C}_{ij}$, indicated by $\mathbf{C}_{ijl}$, denotes the water usage of device i at the l^{th} interval in the j^{th} day, where $l = 1, \cdots, T$. The aggregated water consumption over all devices is $\bar{\mathbf{C}} = \sum_{i=1}^{k} \mathbf{C}_i$. During the course of training, water consumption of individual fixture is supposed to be available, i.e., $\mathbf{C}_1, \mathbf{C}_2, \cdots, \mathbf{C}_k$, while at the testing phase, only the aggregated consumption $\bar{\mathbf{C}}$ could be used for disaggregation. Under the context of sparse coding for source separation [Schmidt and Olsson, 2006; Schmidt *et al.*, 2007], water consumption of the i^{th} fixture is approximated by $\mathbf{C}_i \approx \mathbf{B}_i\mathbf{A}_i$, where $\mathbf{B}_i \in \mathbb{R}^{T\times N}$ denotes the dictionary for fixture i, and $\mathbf{A}_i \in \mathbb{R}^{N\times M}$ is the activations of dictionary $\mathbf{B}_i$. N is the number of basis functions in dictionary $\mathbf{B}_i$, which can be activated by $\mathbf{A}_i$ for the reconstruction of $\mathbf{C}_i$ [Olshausen and Field, 1996].

Discriminative Disaggregation Sparse Coding Discriminative Disaggregation Sparse Coding (DDSC) was proposed in [Kolter *et al.*, 2010] and designed for energy disaggregation. With the assumption that the disaggregation dictionary is not necessary the same as that for reconstruction, they define the augmented regularized disaggregation error as the objective function

$$\begin{aligned} &\tilde{E} = \sum_{i=1}^{k} \left(\frac{1}{2} \left\| \mathbf{C}_i - \mathbf{B}_i \hat{\mathbf{A}}_i \right\|_F^2 + \lambda \sum_{p,q} \left(\hat{\mathbf{A}}_i \right)_{pq} \right) \\ &\text{subject to:} \\ &\hat{\mathbf{A}}_{1:k} = \underset{\mathbf{A}_{1:k} \geq 0}{\operatorname{argmin}} \left\| \bar{\mathbf{C}} - \tilde{\mathbf{B}}_{1:k} (\mathbf{A}_{1:k}^T)^T \right\|_F^2 + \lambda \sum_{i,p,q} (\mathbf{A}_i)_{pq} \end{aligned} \tag{1}$$

Where $\|\mathbf{X}\|_F \equiv \left(\sum_{p,q} (\mathbf{X})_{pq} \right)^{1/2}$ denotes the Frobenius norm, $\lambda \in \mathbb{R}_+$ is a regularization parameter, $\mathbf{X}_{1:k}$ is the shorthand for $[\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_k]$. In Equation (1), $\mathbf{B}_{1:k}$ represents the reconstruction bases, which are learned based on sparse coding model. $\tilde{\mathbf{B}}_{1:k}$ denotes the discriminative bases which intend to reduce the difference between $\hat{\mathbf{A}}_{1:k}$ and $\mathbf{A}^{\star}_{1:k}$, where $\mathbf{A}^{\star}_{1:k}$ is the activations learned by sparse coding. A structured prediction based algorithm was applied for optimization by iteratively updating $\hat{\mathbf{A}}_{1:k}$ and $\tilde{\mathbf{B}}_{1:k}$.

However, DDSC might fail to identify disaggregation dictionary when training bases $\tilde{\mathbf{B}}_{1:k}$ with a large value of k and limited label data. Moreover, when the number of fixtures increases, it's of high probability that consumption patterns of devices are similar with each other. Consequently, the disaggregation performance of DDSC will decrease significantly.

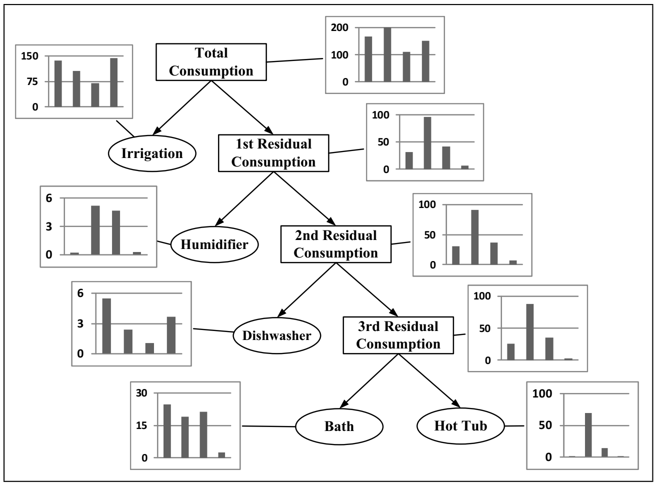

Figure 1: Illustration of the recursive disaggregation architecture.

Since the size of $\tilde{\mathbf{B}}_{1:k}$ increases with k, the vast amount of learnable parameters will lead to many local optima and high time consumption for iterative update. In addition, due to high cost and invasion of privacy concerns, only a small quantity of label data can be collected and available for training. This might result in poor efficacy of DDSC since $\tilde{\mathbf{B}}_{1:k}$ is large.

3 Deep Sparse Coding based Recursive Disaggregation Model

In this section, we design a decomposition structure to overcome the limitations of DDSC and propose the algorithm *deep sparse coding* to perform disaggregation task.

3.1 Design of the Recursive Decomposition Structure

Representation of the structure or data has been considered as a significant factor for the success of machine learning [Bengio *et al.*, 2012], which also plays an important role in the disaggregation task. For example, suppose there are three parallel devices: two similar *Showers* (denoted as SH_1 and SH_2) along with one *Clothes Washer* (denoted as CW). If we disaggregate these three activities simultaneously, it's difficult to discriminate the two similar *Showers*. On the other hand, if we consider a different representation: SH_1 along with the combination of SH_2 and CW, then we can first derive SH_1 from the total consumption, and then separate SH_2 from the combination. Since the consumption patterns of the combination of one *Shower* and one *Clothes Washer* is usually quite different from those of one *Shower*, it's rather easier to achieve a high disaggregation performance based on such a disaggregation structure.

This motivates us to conduct an in-depth study on the representation of decomposition structure, and design a recursive architecture to conduct the disaggregation task. An illustration is shown in Figure 1: The disaggregation process is conducted in a recursive manner to separate one device from the remainders at one layer until all devices are examined. Considering there are five water fixtures, *Irrigation* is firstly extracted while the remainders are combined together and regarded as an artificial device. Then with the usage of the 1st residual consumption, *Humidifier* is separated from the other devices. Similarly, *Dishwasher*, *Bath*, and *Hot Tub* could be identified. Based on the designed structure, it's clear that within each step, only $2 \times N$ basis functions are required to decompose the aggregated consumption into one single device and the remainders. Compared with DDSC, it is capable of fully utilizing the limited label data, since $k-1$ combinations of the data are inherently incorporated in the disaggregation process. To formalize the representation of the disaggregation architecture, we introduce the definition *sequential set*,

Definition 1. Suppose set S contains k elements, i.e., $S = \{1, 2, \cdots, k\}$. Then its *sequential set* $\mathbf{S}_i$, $(1 \le i \le k)$ is defined as: $\mathbf{S}_i = \{\mathbf{S}_i^{(1)}, \mathbf{S}_i^{(2)}\}$, where

$$\mathbf{S}_i^{(1)} = \begin{cases} e \in S & , i = 1 \\ e \in \mathbf{S}_{i-1}^{(2)} & , 2 \le i \le k \end{cases},$$

and $\mathbf{S}_i^{(2)} = S \setminus \bigcup_{j=1}^{i} \left\{ \mathbf{S}_j^{(1)} \right\}, 1 \le i \le k$.

The reason we call $\mathbf{S}_i$ for $1 \le i \le k$ as the *sequential set* is that it inherently sorts set S as $\mathbf{S}_1^{(1)}, \mathbf{S}_2^{(1)}, \cdots, \mathbf{S}_k^{(1)}$. For example, Figure 1 shows one *sequential set* for the devices: $\mathbf{S}_1^{(1)} = s_{\textit{Irrigation}}$, $\mathbf{S}_1^{(2)} = \{s_{\textit{Humidifier}}, s_{\textit{Dishwasher}}, s_{\textit{Bath}}, s_{\textit{Hot Tub}}\}$, $\mathbf{S}_2^{(1)} = s_{\textit{Humidifier}}$, $\mathbf{S}_2^{(2)} = \{s_{\textit{Dishwasher}}, s_{\textit{Bath}}, s_{\textit{Hot Tub}}\}$, $\mathbf{S}_3^{(1)} = s_{\textit{Dishwasher}}$, $\mathbf{S}_3^{(2)} = \{s_{\textit{Bath}}, s_{\textit{Hot Tub}}\}$, $\mathbf{S}_4^{(1)} = s_{\textit{Bath}}$, $\mathbf{S}_4^{(2)} = \{s_{\textit{Hot Tub}}\}$, $\mathbf{S}_5^{(1)} = s_{\textit{Hot Tub}}$, $\mathbf{S}_5^{(2)} = \emptyset$, where $s_\star$ is the order number for the corresponding device. The inherent sorting result is *Irrigation* $\rightarrow$ *Humidifier* $\rightarrow$ *Dishwasher* $\rightarrow$ *Bath* $\rightarrow$ *Hot Tub*, which corresponds to the disaggregation architecture we need to learn.

3.2 Disaggregation Structure Learning via Deep Sparse Coding

Given a set of k devices, we intend to identify a *sequential set* to reveal the optimal disaggregation order, which can maximize the disaggregation performance. Mathematically, given $S = \{1, 2, \cdots, k\}$, we have the following optimization problem,

$$\begin{aligned} \text{Minimize} \quad \mathcal{E} = \sum_{i=1}^{k} \sum_{j=1}^{2} \Bigg[\frac{1}{2} \left\| \mathbf{C}_{\mathbf{S}_i^{(j)}} - \mathbf{B}_{\mathbf{S}_i^{(j)}} \hat{\mathbf{A}}_{\mathbf{S}_i^{(j)}} \right\|_F^2 \\ + \lambda \sum_{p,q} \left(\hat{\mathbf{A}}_{\mathbf{S}_i^{(j)}} \right)_{pq} \Bigg] \end{aligned} \tag{2}$$

subject to:

$$\mathbf{S}_i^{(1)} \in S, \quad \mathbf{S}_i^{(2)} = S \setminus \bigcup_{j=1}^{i} \left\{ \mathbf{S}_j^{(1)} \right\}, \quad (1 \le i \le k). \tag{3}$$

$$\hat{\mathbf{A}}_{\mathbf{S}_1:\mathbf{S}_k} = \operatorname*{argmin}_{\mathbf{A}_{\mathbf{S}_1:\mathbf{S}_k}\geq 0} \sum_{i=1}^{k}\left[\frac{1}{2}\left\|\bar{\mathbf{C}}_{\mathbf{S}_i} - \mathbb{B}_{\mathbf{S}_i}\mathbf{A}_{\mathbf{S}_i}\right\|_F^2 + \lambda\sum_{p,q}(\mathbf{A}_{\mathbf{S}_i})_{pq}\right] \tag{4}$$

$$\mathbf{A}_{\mathbf{S}_i^{(j)}} = \operatorname*{argmin}_{\mathbf{A}\geq 0}\left[\frac{1}{2}\left\|\mathbf{C}_{\mathbf{S}_i^{(j)}} - \mathbf{B}_{\mathbf{S}_i^{(j)}}\mathbf{A}\right\|_F^2 + \lambda\sum_{p,q}(\mathbf{A})_{pq}\right] \tag{5}$$

$$\mathbf{B}_{\mathbf{S}_i^{(j)}} = \operatorname*{argmin}_{\mathbf{B}\geq 0, \|\mathbf{b}^j\|_2\leq 1}\frac{1}{2}\left\|\mathbf{C}_{\mathbf{S}_i^{(j)}} - \mathbf{B}\mathbf{A}_{\mathbf{S}_i^{(j)}}\right\|_F^2 \tag{6}$$

$$\mathbb{B}_{\mathbf{S}_1:\mathbf{S}_k} = \operatorname*{argmin}_{\mathbb{B}_{\mathbf{S}_1:\mathbf{S}_k}\geq 0, \|\mathbf{b}^j\|_2\leq 1}\sum_{i=1}^{k}\frac{1}{2}\left\|\mathbf{A}_{\mathbf{S}_i} - \hat{\mathbf{A}}_{\mathbf{S}_i}\right\|_F^2 \tag{7}$$

Where $\mathcal{E}$ denotes the *recursive disaggregation error*, $\mathbf{C}_{\mathbf{S}_i^{(1)}}$ denotes the consumption of device $\mathbf{S}_i^{(1)}$, $\mathbf{C}_{\mathbf{S}_i^{(2)}} = \sum_{l=1}^{|\mathbf{S}_i^{(2)}|}\mathbf{C}_{(\mathbf{S}_i^{(2)})_l}$ $\left((\mathbf{S}_i^{(2)})_l\right.$ is the l^{th} element in $\left.\mathbf{S}_i^{(2)}\right)$ represents the i^{th} residual consumption, $\bar{\mathbf{C}}_{\mathbf{S}_i} = \mathbf{C}_{\mathbf{S}_i^{(1)}} + \mathbf{C}_{\mathbf{S}_i^{(2)}}$ is the total consumption at the i^{th} layer; $\mathbb{B}_{\mathbf{S}_i} = \left[\mathbb{B}_{\mathbf{S}_i^{(1)}}, \mathbb{B}_{\mathbf{S}_i^{(2)}}\right]$ denotes the discriminative dictionary which is learned by minimizing the difference between $\hat{\mathbf{A}}_{\mathbf{S}_i}$ and $\mathbf{A}_{\mathbf{S}_i}$, $\mathbf{B}_{\mathbf{S}_i^{(j)}}$ is the reconstruction dictionary, $\mathbf{A}_{\mathbf{S}_i^{(j)}}$ is the activations learned by sparse coding, $\hat{\mathbf{A}}_{\mathbf{S}_i^{(j)}}$ is the estimated activations for discrimination, $\mathbf{A}_{\mathbf{S}_i} = \begin{bmatrix}\mathbf{A}_{\mathbf{S}_i^{(1)}} \\ \mathbf{A}_{\mathbf{S}_i^{(2)}}\end{bmatrix}$, and $\mathbf{X}_{\mathbf{S}_1:\mathbf{S}_k} = \{\mathbf{X}_{\mathbf{S}_1}, \cdots, \mathbf{X}_{\mathbf{S}_k}\}$.

This optimization problem is a Non-Convex Mixed-Integer Nonlinear Programming (Non-Convex MINLP) problem, and it's NP-hard in general [Garey and Johnson, 1979]. Specifically, the time complexity of this problem increases exponentially with the number of involved devices, which will cause a prohibitively high time complexity for a large-sized S.

This motivates us to design the algorithm *deep sparse coding* to iteratively solve this problem based on the assumption that the choice of $\mathbf{S}_i^{(1)} (1 \leq i \leq k)$ can be learned sequentially. We introduce the objective function of *deep sparse coding*, which targets for the optimization with each layer,

$$\text{Minimize} \quad \mathcal{E}_i = \sum_{j=1}^{2}\left[\frac{1}{2}\left\|\mathbf{C}_{\mathbf{S}_i^{(j)}} - \mathbf{B}_{\mathbf{S}_i^{(j)}}\hat{\mathbf{A}}_{\mathbf{S}_i^{(j)}}\right\|_F^2 + \lambda\sum_{p,q}\left(\hat{\mathbf{A}}_{\mathbf{S}_i^{(j)}}\right)_{pq}\right] \tag{8}$$

The constraints for the above objective are similar with those formulated by Equations (3), (4), (5), (6), (7), but here we only consider those factors which affect the current layer i. Through the relaxation, in terms of the number of involved devices, the time complexity of the problem decreases from $O(k!)$ to $O(k^2)$. The details of *deep sparse coding* algorithm are shown in Algorithm 1. We use Procedure RDSC to identify the device that needs to be disaggregated at each layer.

Algorithm 1 Deep sparse coding

Input: $\mathbf{C}_i$: consumption for device i; λ : regularization parameter; α : gradient step learning rate; device set $S = \{1, \cdots, k\}$.
Output: $\mathbf{S}_i$: layer-wise optimal sequential set of S; $\mathbb{B}_{\mathbf{S}_i^{(j)}}$: discriminative dictionary; $\mathbf{B}_{\mathbf{S}_i^{(j)}}$: reconstruction dictionary.

Initialization.
foreach i in S
  $(\mathbf{A}_i, \mathbf{B}_i) \leftarrow \text{SC}(\mathbf{C}_i)$.
$m \leftarrow k$. $l \leftarrow 1$.
RDSC (l, S, $\mathbf{C}_{1:m}$, $\mathbf{B}_{1:m}$, $\mathbf{A}_{1:m}$).

Procedure RDSC(l, S, $\mathbf{C}_{1:m}$, $\mathbf{B}_{1:m}$, $\mathbf{A}_{1:m}$)
if ($S == \emptyset$) **Then** RETURN.
foreach i in S
  $\mathbf{C}_{R_i} \leftarrow \sum_{j\in S, j\neq i}\mathbf{C}_j$. $\bar{\mathbf{C}} \leftarrow \mathbf{C}_i + \mathbf{C}_{R_i}$.
  $(\mathbf{A}_{R_i}, \mathbf{B}_{R_i}) \leftarrow \text{SC}(\mathbf{C}_{R_i})$.
  $(\mathbb{B}_i, \mathbb{B}_{R_i}, \hat{\mathbf{A}}_i, \hat{\mathbf{A}}_{R_i}) \leftarrow \text{DDSC}(\bar{\mathbf{C}}, [\mathbf{B}_i, \mathbf{B}_{R_i}], [\mathbf{A}_i, \mathbf{A}_{R_i}])$.
$\mathbf{S}_l^{(1)} \leftarrow \operatorname*{argmin}_{i\in S}\mathcal{E}_i$. $\mathbf{S}_l^{(2)} \leftarrow S \setminus \{\mathbf{S}_l^{(1)}\}$. $\mathbf{S}_l \leftarrow \{\mathbf{S}_l^{(1)}, \mathbf{S}_l^{(2)}\}$.
$c \leftarrow \mathbf{S}_l^{(1)}$. $\mathbb{B}_{\mathbf{S}_l} \leftarrow \{\mathbb{B}_c, \mathbb{B}_{R_c}\}$. $\mathbf{B}_{\mathbf{S}_l} \leftarrow \{\mathbf{B}_c, \mathbf{B}_{R_c}\}$.
RDSC($l+1, \mathbf{S}_l^{(2)}, \mathbf{C}_{1:m}\setminus\mathbf{C}_{\mathbf{S}_l^{(1)}}, \mathbf{B}_{1:m}\setminus\mathbf{B}_{\mathbf{S}_l^{(1)}}, \mathbf{A}_{1:m}\setminus\mathbf{A}_{\mathbf{S}_l^{(1)}}$).

Procedure SC($\mathbf{C}_i$)
Initialize $\mathbf{A}_i$, $\mathbf{B}_i$ with non-negative values.
Iterate until convergence:
  $\mathbf{A}_i \leftarrow \operatorname*{argmin}_{\mathbf{A}\geq 0}\frac{1}{2}\|\mathbf{C}_i - \mathbf{B}_i\mathbf{A}\|_F^2 + \lambda\sum_{p,q}(\mathbf{A})_{p,q}$.
  $\mathbf{B}_i \leftarrow \operatorname*{argmin}_{\mathbf{B}\geq 0, \|\mathbf{b}^{(j)}\|_2\leq 1}\frac{1}{2}\|\mathbf{C}_i - \mathbf{B}\mathbf{A}_i\|_F^2$.
RETURN ($\mathbf{A}_i$, $\mathbf{B}_i$).

Procedure DDSC($\bar{\mathbf{C}}$, $[\mathbf{B}_1, \mathbf{B}_2]$, $[\mathbf{A}_1, \mathbf{A}_2]$)
$\mathbf{A}_{1:2}^{\star} \leftarrow \mathbf{A}_{1:2}$. $\mathbb{B}_{1:2} \leftarrow \mathbf{B}_{1:2}$. $\mathbb{B} \leftarrow [\mathbb{B}_1, \mathbb{B}_2]$.
Iterate until convergence:
  $\hat{\mathbf{A}}_{1:2} \leftarrow \operatorname*{argmin}_{\mathbf{A}_{1:2}\geq 0}\left[\frac{1}{2}\left\|\bar{\mathbf{C}} - \mathbb{B}_{1:2}\mathbf{A}_{1:2}\right\|_F^2 + \lambda\sum_{p,q}(\mathbf{A}_{1:2})_{pq}\right]$.
  $\mathbb{B} \leftarrow \left[\mathbb{B} - \alpha\left(\left(\bar{\mathbf{C}} - \mathbb{B}\hat{\mathbf{A}}\right)\hat{\mathbf{A}}^T - \left(\bar{\mathbf{C}} - \mathbb{B}\mathbf{A}^{\star}\right)(\mathbf{A}^{\star})^T\right)\right]_+$.
  for all i, j, $\mathbf{b}_i^{(j)} \leftarrow \mathbf{b}_i^{(j)}/\|\mathbf{b}_i^{(j)}\|_2$.
RETURN ($\mathbb{B}_1, \mathbb{B}_2, \hat{\mathbf{A}}_1, \hat{\mathbf{A}}_2$).

RDSC can fully utilize Procedure DDSC in different layers to discriminatively optimize the bases $\mathbb{B}$ by reducing the difference between reconstruction and discrimination activations. Consequently, we can learn the sequential set $\mathbf{S}_i$ of S, which specifies the disaggregation order. Both discriminative and reconstruction bases can also be learned for disaggregation.

Estimation of Device Level Consumption Based on the structure and parameters learned by deep learning algorithm (Algorithm 1), we can estimate device level consumption. The details of the disaggregation process are shown in Algorithm 2.

Algorithm 2 Device Level Consumption Estimation

Input: $\bar{\mathbf{C}}$: aggregated consumption; $\mathbf{S}_i$: sequential set of S; $\mathbb{B}_{\mathbf{S}_i^{(j)}}$: discriminative dictionary; $\mathbf{B}_{\mathbf{S}_i^{(j)}}$: reconstruction dictionary.

Output: $\hat{\mathbf{C}}_i$: estimated consumption for device i.

1: Initialization.
2: **for** $i \leftarrow 1$ to k
3: $\quad \hat{\mathbb{A}}_{1:2} \leftarrow \underset{\mathbb{A}_{1:2} \geq 0}{\operatorname{argmin}} \left[\frac{1}{2} \left\| \bar{\mathbf{C}} - \mathbb{B}_{\mathbf{S}_i} \mathbf{A}_{1:2} \right\|_F^2 + \lambda \sum_{p,q} (\mathbf{A}_{1:2})_{pq} \right]$.
4: $\quad \hat{\mathbf{C}}_{\mathbf{S}_i^{(1)}} \leftarrow \mathbf{B}_{\mathbf{S}_i^{(1)}} \hat{\mathbb{A}}_1$.
5: $\quad \bar{\mathbf{C}} \leftarrow \mathbf{B}_{\mathbf{S}_i^{(2)}} \hat{\mathbb{A}}_2$.

4 Experimental Evaluation

In this section, we conducted a large scale water disaggregation task based on a real-world data set.

4.1 Experiment Setup

Data Set: We conducted water disaggregation with a data set collected by Aquacraft [Mayer *et al.*, 1999]. The data set contains 1,959,817 water use events recorded during a two-year study from 1,188 residents across 12 study sites (e.g., San Diego, Denver, Eugene, and Seattle) located in six distinct regions of North America. Since the widely deployed smart meters are constrained to report at a low sample rate (typically $1/900$ Hz), for generalization purpose, we converted the event records into interval based time series with a $1/900$ Hz sample rate. In the course of experiments, we performed training/testing with respect to each household: 70% of the data for training, 30% of the data for testing.

Benchmark Methods: Two baselines were considered for comparison. One is DDSC model proposed in [Kolter *et al.*, 2010], which was designed for energy disaggregation. The other one is FHMM based disaggregation approach presented in [Kim *et al.*, 2011; Kolter and Johnson, 2011], which was capable of disaggregating power data. Since the features required for CFHMM in [Kim *et al.*, 2011] are not available in the scenario of water disaggregation, we naturally used FHMM instead of CFHMM.

Evaluation Metrics: Both device and whole home level evaluation metrics were used to measure the performance. The whole home disaggregation performance was evaluated with accuracy and Normalized Disaggregation Error (NDE),

$$\text{Accuracy} = \frac{\sum_{i,q} \min \left(\sum_p (\mathbf{C}_i)_{pq}, \sum_p (\hat{\mathbf{C}}_i)_{pq} \right)}{\sum_{p,q} (\bar{\mathbf{C}})_{pq}} \tag{9}$$

$$\text{NDE} = \sqrt{\frac{\left(\sum_{i,p,q} \left\| (\mathbf{C}_i)_{pq} - (\hat{\mathbf{C}}_i)_{pq} \right\| \right)^2}{\left(\sum_{i,p,q} (\mathbf{C}_i)_{pq} \right)^2}} \tag{10}$$

Where $\hat{\mathbf{C}}_i = \mathbf{B}_i \hat{\mathbf{A}}_i$ is the estimated consumption of device i. Large accuracy indicates that the estimated consumption can cover the true data well; while large NDE means that the estimated consumption is significantly different from the ground truth. In addition, the metrics used to evaluate the device level disaggregation performance include precision, recall, and F-measure. Precision is the fraction of estimated consumption that is correctly separated, recall indicates the fraction of true consumption that is successfully classified, and F-measure is calculated by $2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$.

4.2 Performance Evaluation and Comparison

Device Level Disaggregation Performance: We assessed DSCRDM, DDSC, and FHMM with 12 domestic water fixtures. Based on Algorithm 1, we learned the recursive disaggregation structure and its corresponding disaggregation sequence for these 12 water devices was $11 \rightarrow 4 \rightarrow 6 \rightarrow 2 \rightarrow 9 \rightarrow 7 \rightarrow 5 \rightarrow 10 \rightarrow 3 \rightarrow 8 \rightarrow 12 \rightarrow 1$.

As shown in Table 1, DSCRDM significantly outperformed DDSC and FHMM. Specifically, DSCRDM achieved **48.53%** in average F-measure (the mean value of all devices' F-measure), while DDSC and FHMM only respectively achieved 22.89% and 16.80%. DSCRDM was capable of attaining more than **80%** precision for *Humidifier*, *Treatment*, *Irrigation*, and *Swimming Pool*, and achieving more than **70%** recall for *Faucet* and *Shower*. Interestingly, DDSC outperformed DSCRM in recall with respect to *Dishwasher*, *Humidifier*, *Treatment*, *Hot Tub*, and *Bath*, but DDSC reached a relatively low precision for these devices. The *low precision and high recall phenomenon* was caused by the fact that DDSC failed to identify the discriminative dictionaries, and the estimated consumption was much larger than true consumption, causing good recall and bad precision. Similarly, FHMM was able to achieve 48.55% in recall but only 9.74% in precision.

Household Level Disaggregation Performance: The whole home level accuracy and normalized disaggregation error comparison between DSCRDM, DDSC, and FHMM are shown in Figure 2. DSCRDM significantly outperformed FHMM with respect to accuracy. However, it failed to be more successful than DDSC. Referring to Equation 9, we could conclude that the estimated consumption of DDSC is much larger than the true consumption, which is also in accordance with the *low precision and high recall phenomenon* mentioned above. Since the *low precision and high recall phenomenon* occurred in five water fixtures, it largely distorted the authenticity of accuracy. NDE achieved by DSCRDM was as small as **0.7417**, which was much smaller than those attained by either DDSC or FHMM. Since NDE measures the difference between the estimated and actual consumption, we could infer that DSCRDM better estimated the overall water usage.

Water Conservation with Detailed Feedbacks We applied DSCRDM to estimate the device level consumption and provide water usage information to users for water conservation. Figure 3 shows that the Actual Consumption (Actual Con.) in one home for a short period from 7:15am to 12:15am, along with the Estimated Consumption (Estimated Con.) in the testing set by DSCRDM. In whole home level, DSCRDM was capable of correctly capturing most of the actual consumption in spite of the small estimation error. The figure also illustrates both true and estimated time series of four selected devices of this home. In many cases, the devices

Devices	DSCRDM	DDSC	FHMM
1. Faucet	19.11%	16.12%	9.94%
	70.08%	22.99%	50.85%
	30.03%	18.95%	16.62%
2. Dishwasher	59.85%	6.06%	8.57%
	59.87%	60.31%	22.92%
	59.86%	11.01%	12.47%
3. Toilet	38.16%	30.71%	9.74%
	32.48%	15.55%	48.55%
	35.09%	20.65%	16.23%
4. Humidifier	84.75%	37.50%	3.66%
	51.27%	85.53%	35.32%
	63.89%	52.13%	6.63%
5. Cooler	77.16%	7.43%	74.63%
	48.57%	6.38%	40.92%
	59.61%	6.87%	52.86%
6. Treatment	83.83%	24.15%	6.28%
	28.68%	70.71%	17.55%
	42.74%	36.01%	9.25%
7. Hot Tub	66.28%	11.86%	9.34%
	48.20%	70.77%	21.59%
	55.82%	20.32%	13.04%
8. Shower	75.77%	7.23%	6.17%
	71.36%	43.50%	18.53%
	73.50%	12.40%	9.26%
9. Bath	53.17%	11.98%	7.03%
	43.09%	53.36%	20.83%
	47.60%	19.57%	10.51%
10. Clothes Washer	42.24%	18.68%	9.86%
	50.64%	14.93%	27.52%
	46.06%	16.59%	14.52%
11. Irrigation	81.36%	65.72%	28.77%
	30.53%	26.60%	10.11%
	44.40%	37.87%	14.96%
12. Swimming Pool	84.48%	40.17%	57.40%
	13.81%	15.38%	16.21%
	23.74%	22.24%	**25.28%**

Table 1: Device level performance for DSCRDM, DDSC, and FHMM. Performance is reported as precision recall F-measure (each occupies one line).

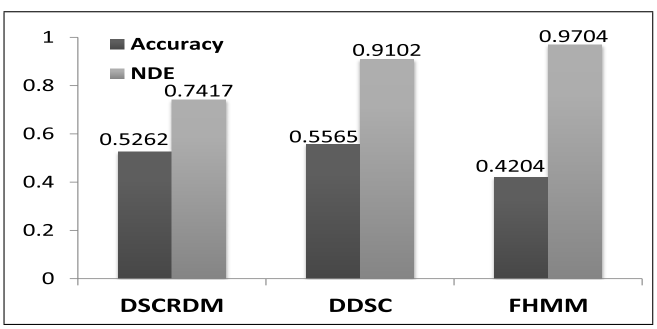

Figure 2: Whole home level performance of DSCRDM, DDSC and FHMM.

Clothes Washer, and *Dishwasher* were estimated accurately. There were also cases DSCRDM could not perform estimation precisely, such as the premature estimation of the spike for *Shower*, and the underestimation of the peak of *Faucet*. Nonetheless, despite some poor estimations, the estimated

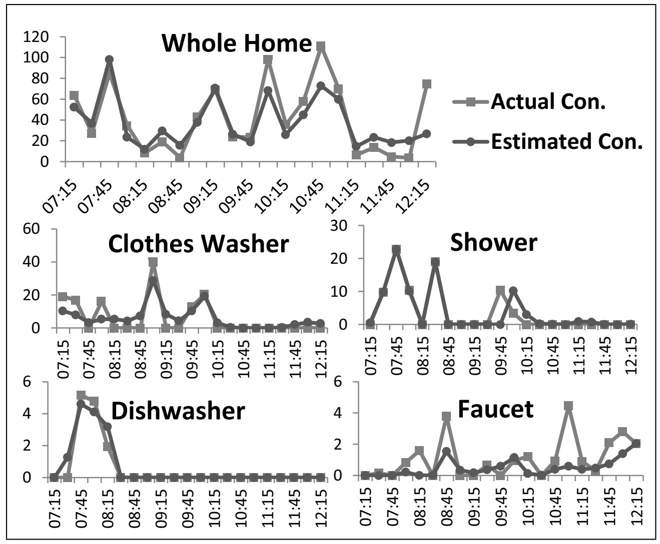

Figure 3: Illustration of the true consumption and estimated consumption by DSCRDM, both in units of Gallons.

water consumption is still quite informative, which could effectively help consumers save a significant amount of water. We investigated the general disaggregation performance of DSCRDM over all devices with the usage pie charts. As shown in Figure 4, DSCRDM could correctly identify the usage percentage information for most devices, although it misclassified some of *Irrigation* as *Faucet*. Such high level usage information could inform users that *Irrigation* occupies more than 70% of the total, indicating that further actions need to be taken, such as using the most water-efficient drip irrigation.

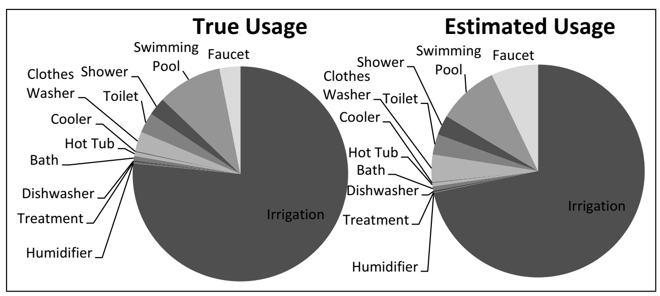

Figure 4: Total water consumption percentages.

5 Conclusion

Water disaggregation, which separates total consumption into individual device consumption, has significant contribution to water conservation. In this paper, we examine the importance of structure learning for disaggregation task and present a machine learning model to automatically discover the decomposition structure. The extensible experiments showed that our algorithm significantly outperformed the benchmark methods. The estimated consumption results by our proposed model were specifically inspected, and detailed feedbacks could be provided to help customers for conservation.

References

[Bengio *et al.*, 2012] Y. Bengio, A. C. Courville, and P. Vincent. Representation learning: A review and new perspectives. *CoRR*, abs/1206.5538, 2012.

[Concern-Worldwide, 2008] Concern-Worldwide. How can we increase the worlds access to clean water? *http://gcc.concernusa.org/media/pdf/Water.pdf*, 2008.

[Fischer, 2008] C. Fischer. Feedback on household electricity consumption: a tool for saving energy? *Energy Efficiency*, 1:79–104, 2008.

[Froehlich *et al.*, 2009] J. E. 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 11th International Conference on Ubiquitous Computing*, pages 235–244, 2009.

[Froehlich *et al.*, 2010] J. Froehlich, L. Findlater, and J. Landay. The design of ecofeedback technology. In *Proceedings of the 28th International Conference on Human factors in computing systems*, Atlanta, GA, USA, 2010.

[Froehlich *et al.*, 2011] 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 event in the home. *Pervasive Computing*, pages 50–69, 2011.

[Garey and Johnson, 1979] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-completeness*, pages 245–248. W.H. Freeman and Company, 1979.

[Houde *et al.*, 2013] S. Houde, A. Todd, A. Sudarshan, J. A. Flora, and K. C. Armel. Real-time feedback and electricity consumption: A field experiment assessing the potential for savings and persistence. *The Energy Journal*, 1(1):87–102, 2013.

[Kim *et al.*, 2011] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han. Unsupervised disaggregation of low frequency power measurements. In *Proceedings of the 2011 SIAM International Conference on Data Mining*, pages 747–758, 2011.

[Kolter and Johnson, 2011] J. Z. Kolter and M. J. Johnson. Redd: A public data set for energy disaggregation research. In *proceedings of the SustKDD workshop on Data Mining Applications in Sustainability*, 2011.

[Kolter *et al.*, 2010] J. Z. Kolter, S. Batra, and A. Y. Ng. Energy disaggregation via discriminative sparse coding. In *Neural Information Processing Systems*, pages 1153–1161, 2010.

[Larson *et al.*, 2012] 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*, 8:82–102, 2012.

[Mayer *et al.*, 1999] P. Mayer, W. DeOreo, E. M. Opitz, J. C. Kiefer, W. Y. Davis, B. Dziegielewski, and J. O. Nelson. *Residential End Uses of Water*. AWWA Research Foundation and American Water Works Association, 1999.

[Olshausen and Field, 1996] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, pages 607–609, 1996.

[Parson *et al.*, 2011] O. Parson, S. Ghosh, M. Weal, and A. Rogers. Using hidden markov models for iterative non-intrusive appliance monitoring. In *Neural Information Processing Systems workshop on Machine Learning for Sustainability*, 2011.

[Schmidt and Olsson, 2006] M. N. Schmidt and R. K. Olsson. Single-channel speech separation using sparse non-negative matrix factorization. In *International Conference on Spoken Language Processing*, pages 2614–2617, 2006.

[Schmidt *et al.*, 2007] M. N. Schmidt, J. Larsen, and F. Hsiao. Wind noise reduction using non-negative sparse coding. In *IEEE Workshop on Machine Learning for Signal Processing*, pages 431–436, 2007.

[Srinivasan *et al.*, 2011] V. Srinivasan, J. Stankovic, and K. Whitehouse. Watersense: Water flow disaggregation using motion sensors. In *The 3rd ACM Workshop On Embedded Sensing Systems For Energy-Efficiency In Buildings*, 2011.

[UNICER, 2012] UNICER. Water, sanitation and hygiene. *http://www.unicef.org/wash/*, 2012.

[Wang *et al.*, 2012] B. Wang, F. Chen, H. Dong, A. P. Boedihardjo, and C.T. Lu. Signal disaggregation via sparse coding with featured discriminative dictionary. In *Proceedings of the 12th International Conference on Data Mining*, volume 12, pages 1134–1139, 2012.

[WHO, 2012] WHO. Global danalysis and assessment of sanitation and drinking-water. *http://whqlibdoc.who.int /publications/2012/9789241503365_eng.pdf*, 2012.