# On Nonintrusive Monitoring of Electrical Appliance Load Via Restricted Boltzmann Machine with Temporal Reservoir

Megumi Fujita[1][a], Yu Fujimoto[2][b] and Yasuhiro Hayashi[1][c]
[1]*Graduate School of Advanced Science and Engineering, Waseda University, Tokyo, Japan*
[2]*Advanced Collaborative Research Organization for Smart Society, Waseda University, Tokyo, Japan*

Keywords: Nonintrusive Appliance Load Monitoring, Gaussian-Softmax Restricted Boltzmann Machine, Gaussian Mixture Model, Reservoir Computing.

Abstract: This study proposes a nonintrusive appliance load monitoring framework for estimation of the power consumptions of individual residential appliances by using aggregated total consumption based on Gaussian-softmax restricted Boltzmann machine with temporal reservoir. The proposed method is expected to estimate the hidden states of the appliances well by coding the current situation in the nonlinear temporal dynamics of the power consumption in the reservoir units so as to estimate the appliance-wise consumptions well. The accuracy of the proposed framework is evaluated by using real-world power consumption data sets.

## 1 INTRODUCTION

The smart electricity meters are currently being penetrating for various electric power consumers in many parts of the world. These smart meters gather power consumption data with temporally high resolution basically for billing purposes of electric utilities; data collected by smart meters are expected to produce a new social value and play an important role in the concept of the Internet of Things (Abate et al., 2018). However, the current smart meter only has the function of metering total power consumption at the meter installation site, so that many additional sensors will be required for presenting the appliance-wise power consumption to encourage the suppression of nonessential demand behavior. This paper discusses a method for estimation of the individual appliance power consumption by utilizing total power consumption data collected by smart meters without requiring additional sensors.

An estimation task of appliance-wise energy consumption by utilizing information of aggregated load collected at a single sensor is called the *energy disaggregation* or the *nonintrusive appliance load monitoring* (NILM), and has been discussed since the 1990s (Hart, 1992; Leeb et al., 1995; Cole and Albicki, 1998). The most of the early works have been focused on relatively high frequency data larger than 1Hz (Hart, 1992; Leeb et al., 1995; Laughman et al., 2003), and have also focused on the reactive power or the other characteristics (Hart, 1992; Leeb et al., 1995; Cole and Albicki, 1998; Laughman et al., 2003; Lee et al., 2005). Meanwhile, there are several other NILM studies focusing on the relatively low frequency active power data. Powers et al. (1991) has utilized the active power information stored every 15 minutes; they focused on the spike form of a large load and proposed the rule based energy disaggregation framework. Farinaccio and Zmeureanu (1999), and Marceau and Zmeureanu (2000) have utilized the active power stored every 16 seconds and proposed another rule-based approach focusing on statistics on the appliance running duration. In recent decade, the idea of factorial hidden Markov model (FHMM) (Ghahramani and Jordan, 1997) and application of matrix factorization technique (Lee and Seung, 1999) have been also applied for energy disaggregation task (Kim et al., 2011; Kolter et al., 2010; Miyasawa et al., 2019), especially by using relatively low-frequency active power data set. In particular, the former is an attractive approach in the sense that it is possible to derive disaggregation results nearly online using data dynamically observed by a smart meter; however, the characteristics in temporal transition of the appliance load consumption vary greatly depending on the appliance, so the construction of flex-

[a] https://orcid.org/0000-0002-8163-3664
[b] https://orcid.org/0000-0001-8475-0535
[c] https://orcid.org/0000-0002-4009-4430



Fujita, M., Fujimoto, Y. and Hayashi, Y.
On Nonintrusive Monitoring of Electrical Appliance Load Via Restricted Boltzmann Machine with Temporal Reservoir.
DOI: 10.5220/0009164109020909
In *Proceedings of the 12th International Conference on Agents and Artificial Intelligence (ICAART 2020) - Volume 2*, pages 902-909
ISBN: 978-989-758-395-7; ISSN: 2184-433X

ible and appropriate models within the scope of naive Markov process is generally difficult. In recent years, the various artificial neural network schemes also have been widely applied to the NILM task (Kelly and Knottenbelt, 2015; He and Chai, 2016; Bonfigli et al., 2018; Mengistu et al., 2019). However, effective approaches to achieve NILM considering nonlinear temporal behavior of load consumption flexibly is still under discussion. For a comprehensive survey of the NILM framework, we refer the readers to Zeifman and Roth (2011), and Ruano et al. (2019).

In this paper, we propose a framework of the NILM task based on the Gaussian-softmax restricted Boltzmann machines (GSRBMs) (Sohn et al., 2011) trained in a supervised manner for identifying hidden states of the target appliances; in particular, we introduce the concept of *reservoir computing* (Chatzis and Demiris, 2011) as a framework that takes into account the nonlinear characteristics of temporal load consumption inherent to each appliance. We also discuss the proposed framework by experimentally evaluating the accuracy of our proposed method using the real-world data set.

The rest of the paper is organized as follows. In Section 2, we explain the basic idea of NILM approach based on the GSRBM used in this paper. In Section 3, we introduce an idea of reservoir for representation of the temporal behavior of the appliance load. In Section 4, we show the simulation results of the proposed approach on real-world data sets for evaluation. Finally, concluding remarks of the study are given in Section 5.

## 2 NILM BASED ON GSRBM

The appliance load monitoring discussed in this paper is based on the accumulated active power consumption data set collected at a single meter in a household every one minute. We assume that data sets of appliance-wise power consumptions are utilized for learning background states of each appliance in a supervised-manner. Firstly, a basic tool used in our approach, i.e. the GSRBM, is briefly introduced. Then, an approach for NILM based on GSRBMs is described.

### 2.1 GSRBM for Estimation of Hidden States of Appliance Load

Let $v_i \in \mathbb{R}_+$ be the power consumed by the appliance $i$ in a household, and $\boldsymbol{h}_i = \{h_{ij}\} \in \{0,1\}^{J_i}$ indicate the hidden load states of the appliance $i$ where $J_i$ represents the number of possible states of the appliance $i$.

We focus on the following type of restricted Boltzmann machine (RBM) (Smolensky, 1986),

$$p(v_i, \boldsymbol{h}_i; \theta_i) = \frac{1}{Z(\theta_i)} \exp(-E(v_i, \boldsymbol{h}_i; \theta_i)), \tag{1}$$

where $E(\cdot)$ is the energy function of the model with the parameter set $\theta_i = \{b_i^{\mathrm{v}} \in \mathbb{R}, b_{ij}^{\mathrm{h}} \in \mathbb{R}\ (j = 1, \ldots, J_i), w_{ij}^{\mathrm{vh}} \in \mathbb{R}\ (j = 1, \ldots, J_i), \sigma_i \in \mathbb{R}_+\}$, which is given as follows,

$$E(v_i, \boldsymbol{h}_i; \theta_i) = -\sum_j \frac{v_i}{\sigma_i} w_{ij}^{\mathrm{vh}} h_{ij} + \frac{(v_i - b_i^{\mathrm{v}})^2}{2\sigma_i^2} - \sum_j h_{ij} b_{ij}^{\mathrm{h}}, \tag{2}$$

and $Z(\theta_i)$ is the partition function defined as follows,

$$Z(\theta_i) = \int_{\infty}^{\infty} \sum_{\boldsymbol{h}_i} \exp(-E(v, \boldsymbol{h}_i; \theta_i)) dv. \tag{3}$$

Here, we introduce the following constraint to $\boldsymbol{h}_i$,

$$\sum_j h_{ij} \leq 1. \tag{4}$$

Then, the model is known to be the Gaussian RBM (Hinton and Salakhutdinov, 2006) with the softmax hidden units, so-called Gaussian-softmax RBM (GSRBM) (Sohn et al., 2011). The conditional probabilities of $v_i$ is given as follows,

$$\begin{aligned} & p(v_i | \boldsymbol{h}_i; \theta_i) \\ & = \frac{p(v_i, \boldsymbol{h}_i; \theta_i)}{\int_{-\infty}^{\infty} p(v, \boldsymbol{h}_i; \theta_i) dv} \\ & = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left( -\frac{(v_i - b_i^{\mathrm{v}} - \sigma_i \sum_j w_{ij}^{\mathrm{vh}} h_{ij})^2}{\sigma_i^2} \right), \end{aligned} \tag{5}$$

and those of $\boldsymbol{h}_i$ under the condition shown in Eq. (4) is given as follows,

$$\begin{aligned} p(\boldsymbol{h}_i | v_i; \theta_i) &= \frac{p(v_i, \boldsymbol{h}_i; \theta_i)}{\sum_{\boldsymbol{h}'_i} p(v_i, \boldsymbol{h}'_i; \theta_i)} \\ &= \frac{\exp\left\{ \sum_j h_{ij} \left( \frac{v_i}{\sigma_i} w_{ij}^{\mathrm{vh}} + b_{ij}^{\mathrm{h}} \right) \right\}}{1 + \sum_j \exp\left( \frac{v_i}{\sigma_i} w_{ij}^{\mathrm{vh}} + b_{ij}^{\mathrm{h}} \right)}. \end{aligned} \tag{6}$$

Note that the condition $\int_{-\infty}^{\infty} p(v; \theta_i) dv = 1$ leads the marginal probability density function of $v_i$ to the following mixture model (McLachlan and Peel, 2000) of $K = J + 1$ Gaussians with the shared variance parameter $\sigma_i$,

$$
\begin{aligned}
p(v_i;\theta_i) &= \sum_{\boldsymbol{h}_i} p(v_i,\boldsymbol{h}_i;\theta_i) \\
&= \sum_{k=1}^{J+1} \pi_{ik} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(v_i-\mu_{ik})^2}{2\sigma_i^2}\right) \\
&= GMM_K(v_i;\psi),
\end{aligned} \tag{7}
$$

where $\psi = \{\{\pi_{ik}\},\{\mu_{ik}\},\sigma_i\}$ is the transformed parameter set for the Gaussian mixture model,

$$
\pi_{ik} = \begin{cases} \frac{\exp\left(b_i^{\mathrm{v}} w_{ij}^{\mathrm{vh}}/\sigma_i + (w_{ij}^{\mathrm{vh}})^2/2 + b_{ij}^{\mathrm{h}}\right)}{1+\sum_j \exp\left(b_i^{\mathrm{v}} w_{ij}^{\mathrm{vh}}/\sigma_i + (w_{ij}^{\mathrm{vh}})^2/2 + b_{ij}^{\mathrm{h}}\right)} & (k=1,\ldots,J) \\ \frac{1}{1+\sum_j \exp\left(b_i^{\mathrm{v}} w_{ij}^{\mathrm{vh}}/\sigma_i + (w_{ij}^{\mathrm{vh}})^2/2 + b_{ij}^{\mathrm{h}}\right)} & (k=J+1(=K)), \end{cases} \tag{8}
$$

is the weight for the mixture, and

$$
\mu_{ik} = \begin{cases} b_i^{\mathrm{v}} + \sigma_i w_{ik}^{\mathrm{vh}} & (k=1,\ldots,J) \\ b_i^{\mathrm{v}} & (k=J+1(=K)), \end{cases} \tag{9}
$$

is the mean parameter of the $k$-th component Gaussian.

In most of the cases, the parameters in this type of RBMs are derived by using the approximated maximum-likelihood approach by using sampling technique, e.g., contrastive divergence learning (Hinton, 2002; Fischer and Igel, 2014); however, the estimation of the variance parameter $\sigma_i$ of GSRBM by the contrastive divergence itself is known to be a difficult task (Hinton, 2012) and it is also known that the estimation becomes very difficult when estimating the other parameters under the small variance parameter $\sigma_i$ even if it is fixed.

In this study, we utilize the fact of equivalence between the GSRBM and the Gaussian mixture model, which was discussed in Sohn et al. (2011), to avoid the difficulties in the learning process. In our implementation, the parameter $\theta_i$ is estimated according to the following steps using the appliance load data $\{v_i(t); t=1,\ldots,T\}$:

**Step 1:** Estimate mixture of $K$ Gaussians with the shared variance parameter by using the EM algorithm (McLachlan and Krishnan, 1996) to obtain the following,

$$
\hat{\psi}_i = \operatorname*{argmax}_{\psi_i} \sum_t \log GMM_K(v_i(t);\psi_i). \tag{10}
$$

**Step 2:** Iterate Step 1 under various $K$, and select the plausible number of components from the viewpoint of the Bayes information criterion (BIC) (Schwarz, 1978).

**Step 3:** Determine $\sigma_i$ by using the estimated variance parameter of the Gaussian mixture model with the plausible $K$ and substitute the remainder parameters in $\theta_i$ for the GSRBM with $J_i = K-1$ hidden units as follows:

$$
b_i^{\mathrm{v}} = \mu_{iK}, \tag{11}
$$

$$
w_{ij}^{\mathrm{vh}} = \frac{1}{\sigma_i}(\mu_{ij} - b_i^{\mathrm{v}}) \quad (j=1,\ldots,J_i), \tag{12}
$$

$$
b_{ij}^{\mathrm{h}} = \log\frac{\pi_{ij}}{\pi_{iK}} - \frac{(w_{ij}^{\mathrm{vh}})^2}{2} - \frac{1}{\sigma_i} w_{ij}^{\mathrm{vh}} b_i^{\mathrm{v}} \quad (j=1,\ldots,J_i). \tag{13}
$$

## 2.2 NILM via GSRBM

In the NILM process, we focus on the total consumption of the appliances in the household, i.e. $v(t) = \sum_i v_i(t)$, and use the GSRBMs learned beforehand by using appliance-wise time-series load data to estimate $v_i(t)$ for each appliance. Assume that the behaviors of the appliance load are statistically independent, then the optimizer of the following problem leads to the plausible appliance states and the appliance load consumptions,

$$
\begin{aligned}
\{\hat{v}_i(t),\hat{\boldsymbol{h}}_i(t);\forall i\} &= \operatorname*{argmax}_{\{v_i\},\{\boldsymbol{h}_i\}} \prod_i p(v_i,\boldsymbol{h}_i;\theta_i), \\
&= \operatorname*{argmax}_{\{v_i\},\{\boldsymbol{h}_i\}} \sum_i \log p(v_i,\boldsymbol{h}_i;\theta_i), \\
&= \operatorname*{argmin}_{\{v_i\},\{\boldsymbol{h}_i\}} \sum_i E(v_i,\boldsymbol{h}_i;\theta_i),
\end{aligned} \tag{14}
$$

$$
\text{s.t.} \sum_i v_i = v(t). \tag{15}
$$

In this study, we focus on the absolute errors between observed load $v(t)$ and the expected appliance load under possible hidden states $\{\boldsymbol{h}_i\}$ without considering the constraint given in Eq. (15), i.e.,

$$
e(v(t);\{\boldsymbol{h}_i\},\{\boldsymbol{\theta}_i\}) = \left| v(t) - \sum_i \left( b_i^{\mathrm{v}} + \sigma_i \sum_j h_{ij} w_{ij}^{\mathrm{vh}} \right) \right|. \tag{16}
$$

To obtain the suboptimizer of Eq. (14) under the constraint Eq. (15), we extract $M$-smallest sets of $\{\boldsymbol{h}_i\}$ from the viewpoint of $e(v(t);\{\boldsymbol{h}_i\},\{\boldsymbol{\theta}_i\})$ shown in Eq. (16), that is written as $\mathcal{H}_M(v(t))$ in this paper, and searched the best solution from $\mathcal{H}_M(v(t))$ according to Algorithm 1. When the number of possible combinations of hidden states is not large, we can set $M = \prod_i J_i$, so that Algorithm 1 is reduced to the brute force global optimization approach.

In this basic idea, only the snapshot $v(t)$ in a one-dimensional data sequence, $\{v(t); t=1,\ldots,T\}$, Algorithm 1: NILM using GSRBM.

**Input:** $\{v(t), \{\theta_i\}, \mathcal{H}_M(v(t))\}$.
Initialize $\underline{f} \leftarrow \infty$
**for** $\{\check{\boldsymbol{h}}_i\} \in \mathcal{H}_M(v(t))$ **do**
Derive the minimizer of Eq. (17) by using the downhill simplex method (Nelder and Mead, 1965),

$$\{\check{c}_i(t)\} = \operatorname*{argmin}_{\{c_i\}} \sum_i E\left(\frac{\exp(c_i)}{\sum_{i'}\exp(c_{i'})} v(t), \check{\boldsymbol{h}}_i; \theta_i\right). \tag{17}$$

Substitute $\check{v}_i \leftarrow \frac{\exp(\check{c}_i)}{\sum_{i'}\exp(\check{c}_{i'})} v(t)$.
Evaluate $f(\{\check{v}_i\}, \{\check{\boldsymbol{h}}_i\}) = \sum_i E(\check{v}_i, \check{\boldsymbol{h}}_i; \theta_i)$.
**if** $f(\{\check{v}_i\}, \{\check{\boldsymbol{h}}_i\}) < \underline{f}$ **then**
Update $\underline{f} \leftarrow f(\{\check{v}_i\}, \{\check{\boldsymbol{h}}_i\})$, $\{\hat{v}_i(t)\} \leftarrow \{\check{v}_i\}$, and $\{\hat{\boldsymbol{h}}_i(t)\} \leftarrow \{\check{\boldsymbol{h}}_i\}$.
**Output:** $\{\hat{v}_i(t), \hat{\boldsymbol{h}}_i(t); \forall i\}$.

is used for estimating hidden states and load consumptions of the appliances at the timeslice $t$ based on the GSRBM (i.e. equivalent to the Gaussian mixture model), so that the behavior of hidden units in RBMs will not reflect the temporal characteristics of appliance-specific load consumptions. Therefore, in the next section, we propose to introduce the idea of reservoir computing (Chatzis and Demiris, 2011) in order to take into account the dynamics of appliance-specific and nonlinear load consumption in estimation of the hidden states for each appliance, and attempt to improve the accuracy of load monitoring.

# 3 TEMPORAL RESERVOIR FOR MODELING STATE TRANSITION

Here, we introduce the idea of reservoir computing (Chatzis and Demiris, 2011) to model nonlinear temporal dynamics in behavior of appliance load. In this approach, a set of sparsely and randomly connected units, so-called the reservoir, is introduced for each appliance model; the weights between connected units in reservoir remains unchanged even during the learning process. The model discussed here can be interpreted as a type of temporal reservoir machine (Schrauwen and Buesing, 2009) using GSRBM, and is expected to provide a functionality to memorize the historical transition of the appliance load by nonlinearly coding into the states of the reservoir units.

Figure 1 shows the architecture of the GSRBM with temporal reservoir for modeling state transition of appliance $i$. Let $\boldsymbol{r}_i = \{r_{il}\} \in (-1,1)^L$ be the set of reservoir units for appliance $i$, $w^{\mathrm{rh}}_{ilj} \in \mathbb{R}$ be the weight

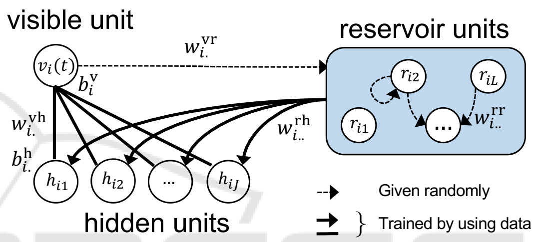


Figure 1: Architecture of the GSRBM with temporal reservoir for modeling state transition of appliance $i$.

from reservoir unit $r_{il}$ to hidden unit $h_{ij}$, $w^{\mathrm{vr}}_{il} \in \mathbb{R}$ be the weight from visible unit $v_i$ to the reservoir unit $r_{il}$, and $W^{\mathrm{rr}}_i = \{w^{\mathrm{rr}}_{ilm}\} \in \mathbb{R}^{L\times L}$ be the matrix containing the weight from the reservoir unit $r_{il}$ to $r_{im}$. We assume that the states of the reservoir units at the timeslice $t+1$, i.e. $\{r_{il}(t+1)\}$, are uniquely determined by the current states $\{r_{il}(t)\}$ and the current appliance load $v_i(t)$ as follows,

$$r_{il}(t+1) = \tanh\left(\sum_m w^{\mathrm{rr}}_{iml} r_{im}(t) + w^{\mathrm{vr}}_{il} v_i(t)\right). \tag{18}$$

Note that the states of these reservoir units are expected to represent the current situation over the historical load transition taking into account the appliance-specific dynamics of power consumption $\{v_i(t)\}$.

We focus on the energy function of the original GSRBM conditioned by the given reservoir units $\boldsymbol{r}_i$ under the architecture shown in Fig. 1 as follows,

$$E(v_i, \boldsymbol{h}_i \mid \boldsymbol{r}_i; \phi_i) = -\sum_j w^{\mathrm{vh}}_{ij} \frac{v_i}{\sigma_i} h_{ij} + \frac{(v_i - b^{\mathrm{v}}_i)^2}{2\sigma_i^2} - \sum_{j,l} w^{\mathrm{rh}}_{ilj} h_{ij} r_{il} - \sum_j h_{ij} b^{\mathrm{h}}_{ij}, \tag{19}$$

where the parameter $\phi_i$ is given as $\phi_i = \{b_i^{\mathrm{v}}, \{b_{ij}^{\mathrm{h}}\}, \{w_{ij}^{\mathrm{vh}}\}, \{w_{ilj}^{\mathrm{rh}}\}, \sigma_i\}$, so that the target conditional probability distribution $p(v_i, \boldsymbol{h}_i \mid \boldsymbol{r}_i; \phi_i)$ is given as follows,

$$p(v_i, \boldsymbol{h}_i \mid \boldsymbol{r}_i; \phi_i) \propto \exp(-E(v_i, \boldsymbol{h}_i \mid \boldsymbol{r}_i; \phi_i)). \quad (20)$$

Under the same constraint given in Eq. (4), the conditional probabilities of $\boldsymbol{h}_i$ is described as the following form,

$$p(\boldsymbol{h}_i \mid v_i, \boldsymbol{r}_i; \phi_i) = \frac{\exp\left\{\sum_j h_{ij}\left(\frac{v_i}{\sigma_i} w_{ij}^{\mathrm{vh}} + \sum_l w_{ilj}^{\mathrm{rh}} r_{il} + b_{ij}^{\mathrm{h}}\right)\right\}}{1 + \sum_j \exp\left(\frac{v_i}{\sigma_i} w_{ij}^{\mathrm{vh}} + \sum_l w_{ilj}^{\mathrm{rh}} r_{il} + b_{ij}^{\mathrm{h}}\right)}, \quad (21)$$

which is affected by the reservoir $\boldsymbol{r}_i$.

In our implementation, the parameter $\phi_i$ in the model shown in Eq. (20) is estimated by the contrastive divergence method using the initial GSRBM parameters derived by the EM algorithm as follows:

**Step 1:** Select the optimal number of hidden units $J_i$ and estimate the GSRBM parameters $\{b_i^v, \{w_{ij}^{\mathrm{vh}}\}, \{b_{ij}^{\mathrm{h}}\}, \sigma_i\}$ as an initial parameter subset.

**Step 2:** Initialize $\{w_{il}^{\mathrm{vr}}\}$, $\{w_{ilj}^{\mathrm{rh}}\}$ and $W_i^{\mathrm{rr}}$: $w_{il}^{\mathrm{vr}}$ is generated from the continuous uniform distribution in the range $[-0.1, 0.1]$; $w_{ilj}^{\mathrm{rh}}$ is initialized by zero; the five percents of the elements randomly selected in the weight matrix $W_i^{\mathrm{rr}}$ are set to have nonzero values also given randomly and scaled so as to hold $\rho(W_i^{\mathrm{rr}}) < 1$ where $\rho(W)$ is the maximum absolute eigenvalue of the matrix $W$.

**Step 3:** Iteratively update the parameters $b_i^{\mathrm{v}}$, $\{b_{ij}^{\mathrm{h}}\}, \{w_{ij}^{\mathrm{vh}}\}, \{w_{ilj}^{\mathrm{rh}}\}$ using the contrastive divergence (CD-1) method (Hinton, 2002);

$$b_i^{\mathrm{v}} \leftarrow b_i^{\mathrm{v}} + \varepsilon \frac{1}{T} \sum_{t=1}^{T} (v_i(t) - \tilde{v}_i(t)), \quad (22)$$

$$b_{ij}^{\mathrm{h}} \leftarrow b_{ij}^{\mathrm{h}} + \varepsilon \frac{1}{T} \sum_{t=1}^{T} \left(h'_{ij}(t) - \tilde{h}_{ij}(t)\right) \ (\forall j), \quad (23)$$

$$w_{ij}^{\mathrm{vh}} \leftarrow w_{ij}^{\mathrm{vh}} + \varepsilon \frac{1}{T} \sum_{t=1}^{T} \left(v_i(t) h'_{ij}(t) - \tilde{v}_i(t) \tilde{h}_{ij}(t)\right) \ (\forall j), \quad (24)$$

$$w_{ilj}^{\mathrm{rh}} \leftarrow w_{ilj}^{\mathrm{rh}} + \varepsilon \frac{1}{T} \sum_{t=1}^{T} r_{il}(t) \left(h'_{ij}(t) - \tilde{h}_{ij}(t)\right) \ (\forall j, l), \quad (25)$$

where $\varepsilon$ is a small positive constant and

$$\boldsymbol{h}'_i(t) = \{h'_{ij}(t)\} \sim p(\boldsymbol{h}_i(t) | v_i(t), \boldsymbol{r}_i(t); \phi_i), \quad (26)$$

$$\tilde{v}_i(t) \sim p(v_i(t) | \boldsymbol{h}'_i(t); \phi_i), \quad (27)$$

$$\tilde{\boldsymbol{h}}_i(t) = \{\tilde{h}_{ij}(t)\} \sim p(\boldsymbol{h}_i | \tilde{v}_i(t), \boldsymbol{r}_i(t); \phi_i). \quad (28)$$

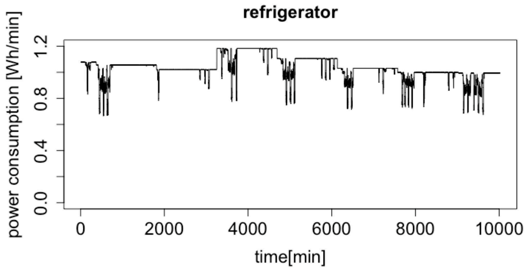
In our proposed scheme, the GSRBMs with temporal reservoir are utilized instead of naive GSRBMs in the NILM process described in Section 2. The estimation of plausible states and load consumptions of the appliances is achieved as

$$\{\hat{v}_i(t), \hat{\boldsymbol{h}}_i(t); \forall i\} = \underset{\{v_i\},\{\boldsymbol{h}_i\}}{\operatorname{argmin}} \sum_i E(v_i, \boldsymbol{h}_i \mid \boldsymbol{r}_i(t); \phi_i), \quad \text{s.t.} \sum_i v_i = v(t), \quad (29)$$

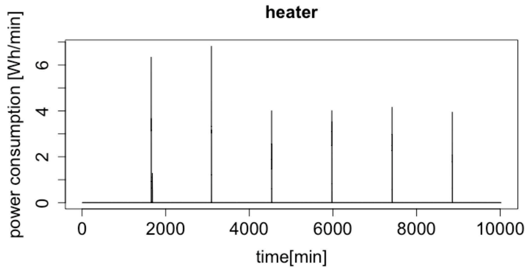
by substituting the energy functions of the models with reservoir, i.e. $\{E(v_i, \boldsymbol{h}_i \mid \boldsymbol{r}_i(t); \phi_i)\}$, for those of naive GBRBMs $\{E(v_i, \boldsymbol{h}_i; \theta_i)\}$ in Eq. (15). Note that in the monitoring process, the reservoir $\{\boldsymbol{r}_{il}(t+1)\}$ is affected by $\hat{v}_i(t)$, which is the disaggregated load consumption of the appliance $i$, and the current states of the reservoir units according to Eq. (18). The optimization problem shown in Eq. (29) can be also searched by Algorithm 1 using the models with reservoir instead of the naive GSRBMs. Since the reservoir has the function of coding the nonlinear dynamics and history of the transition of individual appliance load consumption up to the current timeslice $t$ into $L$ units, this approach is expected to be a way to handle temporal characteristics relatively flexible comparing with a model assuming a simple low-order Markov process of the observed sequence $\{v_i(t)\}$ for each appliance.

# 4 NUMERICAL EXPERIMENTS

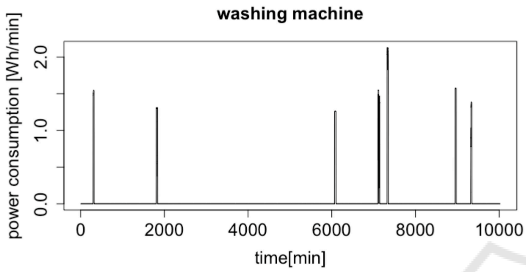
The proposed NILM framework was numerically evaluated from the viewpoint of estimation accuracy using real-world data sets collected at ten households. The data sets were composed of power consumptions of four appliances, i.e. the refrigerator, the washing machine, the heater, and the air conditioners, collected every one minute from July 1st to July 7th in 2017. Figure 2 shows an example of the appliance-wise power consumptions used in the experiments. In the experimental setup, the NILM approach was evaluated by using the sequence of total power consumptions of these four appliances collected at one of the ten households; each appliance model was learned by using the set of appliance-wise data collected at the rest of nine households. The evaluation was processed for all the ten households in a manner of the cross validation. We especially focus on evaluating the effectiveness of the reservoir units by comparing with the scheme based on the naive GBRBMs introduced in Section 2, and the simple mean prediction algorithm (Kolter and Johnson, 2011) as a benchmark; the



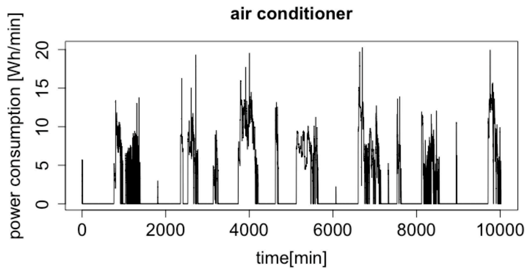

(a) Refrigerator




(c) Heater




(b) Washing machine




(d) Air conditioner

Figure 2: Appliance-wise power consumptions. The input for the NILM task is an aggregated sequence of these consumptions.

latter is given as follows:

$$\hat{v}_i(t) = v(t)\frac{\sum_{t'=1}^{T} v_i^{(\mathrm{lrn})}(t')}{\sum_{t'=1}^{T} v^{(\mathrm{lrn})}(t')}, \qquad (30)$$

where $v_i^{(\mathrm{lrn})}$ and $v^{(\mathrm{lrn})}$ indicate the power consumption of the appliance $i$ and the total power consumption in the learning data set, respectively. The number of hidden units was derived according to the process described in Section 3, and given as 12 for air conditioner, five for washing machine, six for refrigerator, and five for heater, respectively. The number of reservoir units was given as $L = 100$ in this study. The results of the NILM were evaluated from the viewpoint of disaggregation accuracy defined in Kolter and Johnson (2011),

$$\mathrm{ACC} = 1 - \frac{\sum_{t=1}^{T}\sum_i |\hat{v}_i(t) - v_i(t)|}{2\sum_{t=1}^{T} v(t)}, \qquad (31)$$

and that of the appliance-wise percentage mean absolute error[1],

$$\mathrm{PMAE}_i = \frac{100}{T}\sum_{t=1}^{T}\left|\frac{\hat{v}_i(t) - v_i(t)}{v_i(t)}\right|. \qquad (32)$$

Table 1 shows the comparison results of the NILM task for ten households. The result shows that the proposed architecture based on the GSRBM with reservoir is better than the simple mean approach, nor the ordinary GSRBM in most of the cases. The results suggest that the proposed architecture provides a promising NILM framework for handling nonlinear temporal dynamics in behavior of appliance load. Figure 3 shows the comparison results of the appliance-wise percentage mean absolute errors. The figure suggests that the proposed GSRBMs drastically reduce the mean absolute errors by comparing with the simple mean prediction results. Particularly, the estimation errors of the proposed framework based on the GSRBM with reservoir tend to be slightly smaller than those of the ordinary GSRBM, though the results imply that other approaches may need to be consid-

Table 1: Comparison of disaggregation accuracy (ACC).

| House ID | Simple mean | GSRBM w/o reservoir | GSRBM w/ reservoir |
|---|---|---|---|
| 1 | 0.6560 | 0.9688 | **0.9691** |
| 2 | 0.6604 | **0.8866** | **0.8866** |
| 3 | 0.6339 | **0.9690** | **0.9690** |
| 4 | 0.7027 | 0.9604 | **0.9608** |
| 5 | 0.5250 | **0.9418** | **0.9418** |
| 6 | 0.6055 | 0.9488 | **0.9490** |
| 7 | 0.6171 | 0.9236 | **0.9239** |
| 8 | 0.5443 | **0.9599** | 0.9598 |
| 9 | 0.6275 | 0.9407 | **0.9423** |
| 10 | 0.6538 | **0.9083** | **0.9083** |
| Average | 0.6226 | 0.9408 | **0.9411** |

[1]Note that the PMAE for the power consumption of home appliances that are used for a short time can be large.

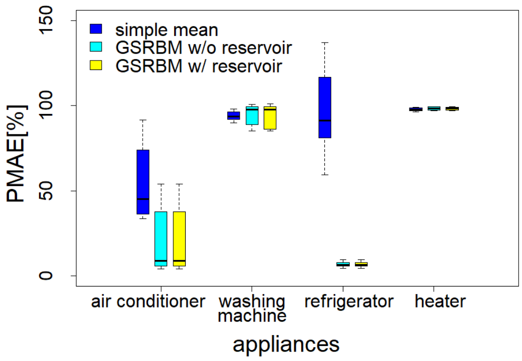



Figure 3: Comparison of appliance-wise percentage mean absolute errors ($\mathrm{PMAE}_i$).

ered to reduce the estimation errors of the appliances that only works occasionally, such as washing machines and heaters.

# 5 CONCLUDING REMARKS

In this study, we focused on a nonintrusive appliance load monitoring scheme based on Gaussian-softmax RBM, and proposed to introduce the temporal reservoir function for the purpose of modeling the nonlinear temporal behavior of the appliance load consumption. The proposed framework was evaluated from the viewpoint of estimation accuracy by numerical experiments. The experimental results show that the introduction of the reservoir tends to improve the estimation accuracy of the appliance-wise energy consumption. Thus, the proposed architecture will be a promising framework for handling nonlinear temporal dynamics in appliance load.

Our experimental results suggested that an explicit mechanism is required to further improve the estimation accuracy for the appliances that only work occasionally, such as washing machines and heaters. Additionally, we also plan to develop a scheme utilizing information stored in the reservoir during the NILM process for the short-term forecast of appliance-wise electricity consumption, though it remains as a future work.

# ACKNOWLEDGEMENTS

A part of this work was supported by the Japan Science and Technology Agency (JST), Core Research for Evolutional Science and Technology (CREST) Program, under Grant JPMJCR15K5.

# REFERENCES


Abate, F., Carratù, M., Liguori, C., Ferro, M., and Paciello, V. (2018). Smart meter for the IoT. In *2018 IEEE International Instrumentation and Measurement Technology Conference: Discovering New Horizons in Instrumentation and Measurement, Proceedings*, pages 1–6.

Bonfigli, R., Felicetti, A., Principi, E., Fagiani, M., Squartini, S., and Piazza, F. (2018). Denoising autoencoders for Non-Intrusive Load Monitoring: Improvements and comparative evaluation. *Energy and Buildings*, 158:1461–1474.

Chatzis, S. P. and Demiris, Y. (2011). Echo state Gaussian process. *IEEE Transactions on Neural Networks*, 22(9):1435–1445.

Cole, A. I. and Albicki, A. (1998). Algorithm for non-Intrusive Identification of Residential Appliances. In *I*, volume III, pages 338–341.

Farinaccio, L. and Zmeureanu, R. (1999). Using a Pattern Recognition Approach to Disaggregate the Total Electricity COnsumption in a House into the Major End-Uses. *Energy and Buildings*, 30:245–259.

Fischer, A. and Igel, C. (2014). Training restricted Boltzmann machines: An introduction. *Pattern Recognition*, 47(1):25–39.

Ghahramani, Z. and Jordan, M. I. (1997). Factorial Hidden Markov Models. *Machine Learning*, 29:245–273.

Hart, G. W. (1992). Nonintrusive Appliance Load Monitoring. *Proceedings of the IEEE*, 80(12):1870–1891.

He, W. and Chai, Y. (2016). An Empirical Study on Energy Disaggregation via Deep Learning. In *2nd International Conference on Artificial Intelligence and Industrial Engineering*, volume 133, pages 338–342.

Hinton, G. E. (2002). Training Products of Experts by Minimizing Contrastive Divergence. *Neural Computation*, 14(8):1771–1800.

Hinton, G. E. (2012). A practical guide to training restricted Boltzmann machines. In *Neural Networks: Tricks of the Trade*, pages 599–619. Springer Berlin Heidelberg.

Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science*, 313:504–507.

Kelly, J. and Knottenbelt, W. (2015). Neural NILM: Deep Neural Networks Applied to Energy Disaggregation. In *BuildSys@SenSys*.

Kim, H., Marwah, M., Arlitt, M., Lyon, G., and Han, J. (2011). Unsupervised Disaggregation of Low Frequency Power Measurements. In *SIAM Conference on Data Mining 2011*, pages 747–758.

Kolter, J. Z., Batra, S., and Ng, A. Y. (2010). Energy Disaggregation via Discriminative Sparse Coding. In *Neural Information Processing Systems*.

Kolter, J. Z. and Johnson, M. J. (2011). REDD: A Public Data Set for Energy Disaggregation Research. In *Workshop on Data Mining Applications in Sustainability*.

Laughman, C., Lee, D., Cox, R., Shaw, S., Leeb, S., Norford, L., and Armstrong, P. (2003). Advanced Nonintrusive Monitoring of Electric Loads. *IEEE Power and Energy Magazine*, pages 56–63.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791.

Lee, K. D., Leeb, S. B., Norford, L. K., Armstrong, P. R., Holloway, J., and Shaw, S. R. (2005). Estimation of Variable-Speed-Drive Power Consumption From Harmonic Content. *IEEE Transactions on Energy Conversion*, 20(3):566–574.

Leeb, S. B., Shaw, S. R., and James L. Kirtley, J. (1995). Transient Event Detection in Spectral Envelope Estimates for Nonintrusive Load Monitoring. *IEEE Transactions on Power Delivery*, 10(3):1200–1210.

Marceau, M. L. and Zmeureanu, R. (2000). Nonintrusive Load Disaggregation Computer Program to Estimate the Energy Consumption of Major End Uses in Residential Buildings. *Energy Conversion and Management*, 41:1389–1403.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley.

McLachlan, G. J. and Krishnan, T. (1996). *The EM Algorithm and Extensions*. Wiley-Interscience.

Mengistu, M. A., Girmay, A. A., Camarda, C., Acquaviva, A., and Patti, E. (2019). A Cloud-Based On-Line Disaggregation Algorithm for Home Appliance Loads. *IEEE Transactions on Smart Grid*, 10(3):3430–3439.

Miyasawa, A., Fujimoto, Y., and Hayashi, Y. (2019). Energy disaggregation based on smart metering data via semi-binary nonnegative matrix factorization. *Energy and Buildings*, 183:547–558.

Nelder, J. and Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313.

Powers, J., Margossian, B., and Smith, B. (1991). Using a Rule-Based Algorithm to Disaggregate End-Use Load Profiles from Premise-Level Data. *IEEE Computer Applications in Power*, 4(2):27–42.

Ruano, A., Hernandez, A., Ureña, J., Ruano, M., and Garcia, J. (2019). NILM techniques for intelligent home energy management and ambient assisted living: A review. *Energies*, 12(11):1–29.

Schrauwen, B. and Buesing, L. (2009). A hierarchy of recurrent networks for speech recognition. *NIPS Workshop on Deep Learning for Speech Recognition and Related Applications*, pages 1–8.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6(2):461–464.

Smolensky, P. (1986). Parallel distributed processing: Explorations in the microstructure of cognition. In *MIT Press, Cambridge, MA, USA*, volume 1, chapter 6, pages 194–281.

Sohn, K., Jung, D. Y., Lee, H., and Hero, A. O. (2011). Efficient learning of sparse, distributed, convolutional feature representations for object recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2643–2650.

Zeifman, M. and Roth, K. (2011). Nonintrusive Appliance Load Monitoring: Review and Outlook. *IEEE Transactions on Consumer Electronics*, 57(1):76–84.