Justifications for the Multi-state Setting

We next discuss the need for a multi-state setting for NILM. The idea is to show that sampling power data from a fine-grained state structure can result in a smaller sample variance, thus attaining a higher probability to reach the mean of real power measurements. We next make some useful assumptions, based on which we present our main theoretical result. Following conventional practice in NILM literature, we assume that an appliance's power consumption follows a normal distribution(Zhang et al. 2018; Mauch and Yang 2016b).

Assumption 1. *For each appliance i, the power consumption of each state s at each time step t is drawn from a normal distribution, i.e., $c_{t,s}^i \sim \mathcal{N}(\mu_s^i, \sigma_s^i)$.*

Recall that the observed power is the expectation on the power of all states, i.e., $y_t^i = \sum_{s=1}^{M^i} p_{t,s}^i c_{t,s}^i$. Due to the independence of power measures of all states and the additivity of normal distributions, we have the following fact.

Fact 1. *Under Assumption 1, the input power y_t^i of an appliance also follows a normal distribution $\mathcal{N}(\mu^i, \sigma^i)$ such that $\mu^i = \sum_{s=1}^{M^i} p_{t,s}^i \mu_s^i$ and $(\sigma^i)^2 = \sum_{s=1}^{M^i} (p_{t,s}^i \sigma_s^i)^2$.*

Since a multi-state model decomposes the total power into a fine-grained state structure, it is natural to propose the following assumption that enforces the variance of the power of each state to not exceed that of the observed total power. This can be seen from the truth depicted in Fig.1: a dishwasher has a steady power level at each state; while, the variance would be increased if we merge any two states into an abstract state.

Assumption 2. *For all $s \in [M^i]$, $\sigma_s^i \leq \sigma^i$.*

We are now ready to present our main result that uses the following notations: $\tilde{y}_t^i$ denotes the sampled power under the single-state assumption; $\bar{y}_t^i = \mathbb{E}_{s\in[M^i]}[\bar{c}_{t,s}^i]$ denotes the sampled power under the multi-state setting which is obtain by averaging the sampled power $\bar{c}_{t,s}^i$ of each state $s \in [M^i]$ (suppose the state information is known a priori). We show in the following theorem that the sampled power data can enjoy a reduced sample variance from the multi-state setting.

Theorem 1. *Suppose we have a sufficiently large number of independent samples. Under Assumptions 1 and 2, the expectations and variances of $\tilde{y}_t^i$ and $\bar{y}_t^i$ satisfy $\mathbb{E}[\bar{y}_t^i] = \mathbb{E}[\tilde{y}_t^i]$ and $\mathbb{D}[\bar{y}_t^i] \leq \mathbb{D}[\tilde{y}_t^i]$ for all $t \in [T]$, where the inequality is strict if $M^i \geq 2$.*

Proof. With a sufficiently large number of independent samples, both $\tilde{y}_t^i$ and $\bar{y}_t^i$ would approach normal distributions, i.e., $\tilde{y}_t^i \sim \mathcal{N}(\mu^i, \sigma^i)$ and $\tilde{y}_t^i = \sum_{s=1}^{M^i} p_{t,s}^i c_{t,s}^i$. Then, according to Fact. 1, $\mathbb{E}[\hat{y}_t^i] = \mathbb{E}[\tilde{y}_t^i]$ follows immediately from the additivity of normal distributions. The reduced variance can be derived from Assumption 2 as follows:

$$\begin{aligned}
\mathbb{D}[\hat{y}_t^i] &= \mathbb{D}\left[\sum\nolimits_{s=1}^{M^i} p_{t,s}^i c_{t,s}^i\right] = \sum\nolimits_{s=1}^{M^i} p_{t,s}^2 \mathbb{D}[c_{t,s}^i] \\
&= \sum\nolimits_{s=1}^{M^i} p_{t,s}^2 \sigma_s^i \leq \sum\nolimits_{s=1}^{M^i} p_{t,s}^2 \sigma^i \leq \left(\sum\nolimits_{s=1}^{M^i} p_{t,s}\right)^2 \sigma^i \\
&= \sigma^i = \mathbb{D}[\tilde{y}_t^i].
\end{aligned}$$

If $M^i \geq 2$, we have $\sum_{s=1}^{M^i} p_{t,s}^2 < \left(\sum_{s=1}^{M^i} p_{t,s}\right)^2$ and hence the inequalities above is strict. □

Corollary 1. *When the appliance has $M^i (M^i \geq 2)$ power states in total, then the power estimation regarding the appliance in a multi-state setting with M^i states would approach the mean value of truth power data with a higher probability than using the single-state setting.*

Proof. Assume $\bar{y}_t^i = \sum_{s=1}^{M^i} p_{t,s}^i c_{t,s}^i$ and $\tilde{y}_t^i$ represent the power estimation utilizing M^i states and one state respectively at each time step. According to Theorem 1, $\bar{y}_t^i$ and $\tilde{y}_t^i$ satisfy normal distribution, which denote as $\bar{y}_t^i \sim \mathcal{N}(\bar{\mu}^i, \bar{\sigma}^i)$ and $\tilde{y}_t^i \sim \mathcal{N}(\tilde{\mu}^i, \tilde{\sigma}^i)$ respectively, then we have $\bar{\mu}^i = \tilde{\mu}^i$, $\bar{\sigma}^i < \tilde{\sigma}^i$. Furthermore for $\forall \xi > 0$:

$$\Pr\left(|\bar{y}_t^i - \bar{\mu}^i| < \xi\right) = \Pr\left(|\frac{\bar{y}_t^i - \bar{\mu}^i}{\bar{\sigma}^i}| < \frac{\xi}{\bar{\sigma}^i}\right) = 2\Phi(\frac{\xi}{\bar{\sigma}^i}) - 1$$

Similarly, we have:

$$\Pr\left(|\tilde{y}_t^i - \tilde{\mu}^i| < \xi\right) = 2\Phi(\frac{\xi}{\tilde{\sigma}^i}) - 1$$

where $\Phi()$ represents the probability meets the standard norm distribution. Since $\bar{\sigma}^i < \tilde{\sigma}^i$, then $\Phi(\frac{\xi}{\bar{\sigma}^i}) < \Phi(\frac{\xi}{\tilde{\sigma}^i})$, further we can get that $\Pr\left(|\bar{y}_t^i - \bar{\mu}^i| < \xi\right) < \Pr\left(|\tilde{y}_t^i - \tilde{\mu}^i| < \xi\right)$. □

Remark 1. *The above corollary ensures that the power estimation of our scheme using multi-state setting can attain a more smaller* MAE *(the main evaluation metric of the scheme, see (performance metrics in Section 5)) error with a higher probability on average than other single state-based schemes.*