Scalable Energy Breakdown Across Regions*

Nipun Batra
University of Virginia
Email: nipunbatra0@gmail.com

Yiling Jia
University of Virginia
Email: yj9xs@virginia.edu

Hongning Wang
University of Virginia
Email: hw5x@virginia.edu

Kamin Whitehouse
University of Virginia
Email: whitehouse@virginia.edu

***Abstract*—Providing an *energy breakdown – energy consumption per appliance*, can help homes save up to 15% energy. Given the vast differences in energy consumption patterns across different regions, existing energy breakdown solutions require instrumentation and model training for each geographical region, which is prohibitively expensive and limits the scalability. In this paper, we propose a novel region independent energy breakdown model via statistical transfer learning. Our key intuition is that the *heterogeneity* in homes and weather across different regions most significantly impacts the energy consumption across regions; and if we can factor out such heterogeneity, we can learn region independent models or the *homogeneous* energy breakdown components for each individual appliance. Thus, the model learnt in one region can be *transferred* to another region. We evaluate our approach on two U.S. cities having distinct weather from a publicly available dataset. We find that our approach gives better energy breakdown estimates requiring the least amount of instrumented homes from the target region, when compared to the state-of-the-art.**

I. Introduction

Various methods for providing an energy breakdown have been studied in the literature. The most intuitive way to get an energy breakdown involves instrumenting each appliance with a sensor. Various sensing systems have been proposed in the past [2], [3]. However, these sensing systems require extensive installation and are thus prohibitively expensive to scale across a large number of homes. In contrast, since the 1980s a novel technique called non-intrusive load monitoring (NILM) has been proposed, which uses statistical techniques to break down the energy measured at the home meter level [4]. However, even NILM would require installing a sensor (such as a smart meter) and would thus cost up to $500 per home, limiting the scalability. Recently, there have been works [5], [6] on providing an energy breakdown without any hardware installation. These approaches can provide an energy breakdown just using the monthly electricity bills and a few homes in the region which already have an energy breakdown. The approaches promise considerable improvement in scalability, but they impose strong assumptions about the problem: the training homes are similar to the testing homes (i.e., all homes are identical and independently distributed). As the set of training homes grows and begins to span multiple climate zones and varied homes (old v/s new, well v/s poorly insulated, studio v/s 3 BHK, etc.), the error will inevitably increase.

In this paper, we present a novel region independent energy breakdown method. Our key insight is that the *heterogeneity* in homes and weather across different regions most significantly impacts the energy consumption across regions; and if we can factor out the weather and homes, we can learn region independent models or the *homogeneous* energy breakdown components for the appliances. As previously shown [5], [7], the heterogeneity across homes (e.g., well-insulated v/s poorly insulated homes) can be captured using a low dimension representation. Intuitively, homes form clusters in this low dimension space. Similarly, the energy dependence of different appliances concerning weather (e.g., the cooling load may be directly proportional to temperature, fridge energy may be season independent, etc.) can also be encoded using a low-dimensional representation. Once we account for the heterogeneity, we can learn *homogeneous* or region-independent factors about the appliances, which capture the interaction between homes and weather. An example of such a factor would be the direct relationship of cooling on home insulation and external temperature. Another example could be the weak relation between fridge energy and home insulation and external temperature.

Our approach thus boils down to learning the heterogeneous home and season/weather factors, referred to as $\mathbf{H}$ and $\mathbf{S}$ respectively; and learning the appliance factor ($\mathbf{A}$) which depends on $\mathbf{H}$ and $\mathbf{S}$. More specifically, we assume the appliance factor $\mathbf{A}$ itself is a three-way tensor (spanning over the number of appliances, and the dimension of $\mathbf{H}$ and $\mathbf{S}$). Intuitively, each appliance factor can be considered as a set of linear combinations of season factors, which form a set of *home bases* (e.g., appliances whose energy usage is sensitive to the change of weather v/s those insensitive). Moreover, each home can thus be characterised as a linear combination over those home bases (e.g., well v/s poorly insulated homes). Therefore, this cross region energy breakdown problem has been naturally formalised as a tensor factorisation problem, where appliance factor $\mathbf{A}$ could be reused across regions. Distinct from standard tensor decomposition solutions (such as PARAFAC [8] and Tucker [9]), which assume the decomposed factors are independent, our solution explicitly encodes the inter-dependency pattern between home factor $\mathbf{H}$ and season factor $\mathbf{S}$ within regions via the appliance factor $\mathbf{A}$, and therefore has the potential to better factor out region dependence.

We evaluate our approach on a publicly available dataset called Dataport [10]. We learn $\mathbf{A}$ factors from 534 homes in Austin and used 40 testing homes from San Diego to evaluate the prediction energy breakdown. Our approach gives better accuracy compared to five baseline approaches, in particular for a low amount of adaptation data required. We see a similar trend when we perform a transfer of $\mathbf{A}$ from 39 homes in San Diego to 40 homes in Austin.

II. Problem Statement

Our aim is to use energy data from a *source* region ($\mathbf{E^{source}}$) and energy data from a small number of *adaptation* homes from the *target* region, to estimate the energy breakdown across homes of the target region. We formally define our *energy tensor* ($\mathbf{E_{M\times N\times T}}$) as a 3-way tensor where the cells contain energy readings of $\mathbf{M}$ homes for $\mathbf{N}$ appliances for $\mathbf{T}$ months. We consider household aggregate energy as one of the "special" appliances. Since aggregate data is easy to collect (via monthly electricity bills), we consider it to be always observed.

Thus, our problem statement can be generally formalised as: given source region energy tensor $\mathbf{E^{source}_{M\times N\times T}}$ and target region energy tensor from a small set of adaptation homes $\mathbf{E^{target}_{Adapt\times N\times T}}$, we want to complete the energy tensor for test homes in target region $\mathbf{E^{target}_{Test\times N\times T}}$. It should be noted that even in the test region, we have aggregate energy always available ($\mathbf{E^{target}_{Test\times 1\times T}}$)

III. Approach: Transferable Tensor Factorisation (TTF)

Our core intuition is that the *heterogeneity* in homes and weather across different regions most significantly impacts the energy consumption across regions; and if we can factor out the weather and homes, we can learn region independent models or the *homogeneous* energy breakdown components. Our energy tensor has three dimensions - home, appliance and weather/season. Since homes and weather

*This paper is a shorter version of the accepted paper at AAAI 2018 with the title: Transferring Decomposed Tensors for Scalable Energy Breakdown across Regions [1] are inherently heterogeneous, we would want the appliance dimension to capture the homogeneity across regions.

Previous work [5] has shown that we can represent the heterogeneity across homes using a low dimension representation. Examples of such low-dimensional representation could be the home insulation, or the number of occupants, the area of the home, etc. Similarly, the energy dependence of different appliances with respect to weather/season (e.g. cooling load may be directly proportional to temperature, fridge energy may be season independent, etc.) can also be encoded using a low-dimensional representation. Thus, on the lines of previous work [5], we can decompose the energy tensor into three factors: home factors (**H**), appliance factors (**A**), and season factors (**S**). We call this tensor decomposition structure as standard tensor factorisation (STF), where, each of **H**, **A** and **S** are independent matrices. In STF, the dimensions of **H**, **A**, **S** are: $M \times r, N \times r$, and $T \times r$, where r is the rank of the energy breakdown tensor.

A fundamental issue with STF is that it does not explicitly factor out the heterogeneity across homes and seasons, as it assumes all three factors are independent from each other. Thus, there is no guarantee that which factor should account for the homogeneity across regions; though it may do a reasonable job in energy breakdown for a single region. To address this problem, we introduce our approach – Transferable Tensor Factorisation (TTF). As mentioned in our assumption, **A** should be learnt as a region-independent factor. Thus, we modify **A** to be another three-way tensor (spanning over the number of appliances, and the dimension of **H** and **S**). Intuitively, each appliance factor can be considered as a set of linear combinations of season factors, which form a set of *home bases* (e.g., appliances whose energy usage is sensitive to the change of weather v/s those insensitive). And each home can thus be characterised as a linear combination of those home bases (e.g., well v/s poorly insulated homes). Therefore, this cross region energy breakdown problem has been naturally formalised as a tensor factorisation problem, where appliance factor **A** could be reused across regions. A caveat of TTF is that it requires more parameters to be learnt compared to STF; which is the trade-off we need to make for learning a region-independent **A** factor.

The key idea of having different heterogeneous and homogeneous factors in energy breakdown is that we only need to learn the heterogeneous components in a new region since the homogeneous components are region independent. This greatly reduces the amount of training data needed from a new region. Based on this idea, our overall procedure for estimating energy breakdown in a target region consists of two steps. In the first step called **Normal learning**, we learn **H**, **A** and **S** from the source region using $\mathbf{E^{source}_{M\times N\times T}}$. In the second step called **Transfer learning**, we reuse the **A** factor learnt from a source domain and only need to learn the **H** and the **S** factors from a few adapt homes. We now describe the two steps.

A. Normal Learning

In normal learning, we learn a model for energy breakdown in a given region. Our idea is to decompose the Energy Breakdown Tensor ($\mathbf{E_{M\times N\times T}}$) into three factors: i) Home factor ($\mathbf{H_{M\times h}}$), ii) Appliance factor ($\mathbf{A_{h\times N\times s}}$), and iii) Season factor ($\mathbf{S_{s\times T}}$), where $\mathbf{h}$ and $\mathbf{s}$ represent the number of home and season factors, respectively.

The normal learning energy tensor decomposition can be represented as:

$$\text{Min } ||\mathbf{E} - \mathbf{HAS}||^2_\mathbf{F} \text{ s.t. } \mathbf{H}, \mathbf{A}, \mathbf{S} \geq \mathbf{0} \quad (1)$$

It should be noted that we enforce non-negativity constraints on each of **H**, **A**, and **S**. This is because energy is a non-negative quantity; hence, each of **H**, **A**, and **S** can only non-negatively contribute to the overall energy.

B. Transfer Learning

For our transfer learning approach, we assume that the appliance factor (**A**) is directly transferable across regions. Thus, for a given target region, we only need to learn the home (H_{target}) and the season factors (S_{target}), which are learnt as per the following optimisation problem,

$$\begin{aligned} \text{Min } ||\mathbf{E_{target}} - \mathbf{H_{target}}\mathbf{A_{source}}\mathbf{S_{target}}||^2_\mathbf{F} \\ \text{s.t. } \mathbf{H_{target}}, \mathbf{S_{target}} \geq \mathbf{0} \end{aligned} \quad (2)$$

If the appliance breakdown of the source domain is representative of the target domain, TTF is expected to work better than normal learning. Otherwise, as more training homes become available, normal learning will eventually do better than transfer learning as it learns an appliance breakdown that is more specific to the target domain.

We can solve the optimisation problem of $\text{Min}||E - HAS||$ using gradient descent. However, as these two optimisation problems are of high dimension and nonconvex, vanilla gradient descent would easily suffer from local minimums. However, variations of gradient descent such as Adagrad [11] have been shown to converge quicker and also use a different learning rate per dimension. Thus, we choose to use Adagrad for updating **H**, **A**, and **S**. We use Autograd [12] for numerical gradient computation. We use the concept of projected gradient descent [13] to ensure that **H**, **A**, and **S** are nonnegative. The following equation shows the procedure of projected gradient descent applied to a variable X. It would apply similarly to update **H**, **A** and **S**.

$$X_{i+1} = \begin{cases} X_i - \delta X_i \times \eta_i, & \text{if } X_i - \delta X_i \times \eta_i \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where X_i, δX_i, η_i represent the value of X, gradient of X, and learning rate at i^{th} iteration.

IV. Evaluation

A. Dataset

In this paper, we use the Dataport [10] dataset for evaluation. While Dataport contains data from various cities in the USA, the maximum instrumentation exists in Austin (Texas) with 534 homes, San Diego (California) with 39 homes, and Boulder (Colorado) with 40 homes. For Boulder, a significant amount of energy consumption comes from appliances which do not have observations in San Diego or Austin. More details about this limitation can be found in the Limitations section. Thus, in this paper, we only consider Austin and San Diego. Energy data from each appliance and the household mains (aggregate) was sampled every minute. Since we plan to evaluate the approach when only monthly electricity data in the form of bills are available, we downsample the data to a month resolution.

Figure 1 shows the energy breakdown across the two regions. It can be seen that Austin has a higher energy footprint compared to San Diego. This can be highly attributed to weather. Austin is a warmer city and thus has a higher cooling energy requirement. Table I shows the proportion of energy contributed by different appliances across the two regions. It is interesting to note that in Austin, HVAC contributes more compared to the fridge; however, in San Diego, the vice-versa trend is visible. The differences in absolute scales in energy consumption, differences in weather, differences in appliance contribution across the two cities, make our problem of transfer learning challenging and interesting.

	HVAC	Fridge	MW	DW	WM	Oven
Austin	0.20	0.09	0.01	0.01	0.01	0.02
San Diego	0.12	0.15	0.02	0.02	0.01	0.03

TABLE I. PROPORTION OF ENERGY CONSUMED BY DIFFERENT APPLIANCES ACROSS AUSTIN AND SAN DIEGO FOR THE YEAR 2014.
* IN THE SUMMER MONTHS, THE HVAC AND THE FRIDGE CAN TOGETHER ACCOUNT FOR $\approx 70\%$ OF AGGREGATE ACROSS BOTH REGIONS.
** WE USE THE FOLLOWING ABBREVIATIONS FOR APPLIANCES: DISHWASHER (DW), MICROWAVE (MW), WASHING MACHINE (WM)

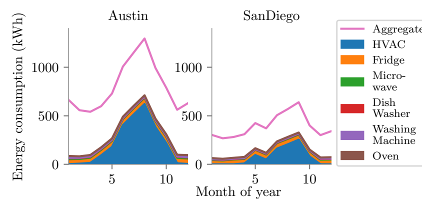

Fig. 1. Energy breakdown across San Diego and Austin across the 12 months in year 2014

B. Baselines

We compare our approach under normal and transfer settings with two approaches. First, we compare against the state-of-the-art approach that leveraged a matrix factorisation (MF) decomposition [5]. The MF approach was only formulated for a single region, which we call as MF under normal settings. In the <MF, Normal> approach, for each appliance w, a matrix $X_w \in R^{M\times(2\times T)}$ is created. The first T columns in this matrix represent the aggregate energy consumption and the last T columns represent the energy consumption of the w^{th} appliance. In this approach, the decomposition for each X_w is done as follows:

$$\begin{aligned} \text{Min } & ||\mathbf{X_w} - \mathbf{Y_w Z_w}||_\mathbf{F}^\mathbf{2} \\ \text{s.t. } & \mathbf{Y_w}, \mathbf{Z_w} \geq \mathbf{0} \end{aligned} \tag{4}$$

where Y_w and Z_w correspond to the latent factors for homes and <appliance, months> respectively. For transfer learning (<MF, Transfer>), we use the Z_w learnt from the source domain and learn Y_w from the target domain. Since MF was shown to be better than the state-of-the-art NILM approaches, we do not compare our work against NILM.

Second, we compare our approach against standard tensor factorisation (STF) introduced earlier in the paper. It should be noted that in TTF we had the following dimensions of **H**, **A**, and **S** respectively: $M \times h, h \times N \times s$, and $s \times T$. However, in STF, the dimensions of **H**, **A**, **S** are: $M \times r, N \times r$, and $T \times r$, where r is the rank of the energy breakdown tensor.

C. Evaluation Metric

Our evaluation metric is based on the prior work [5]. It indicates how close the predicted energy breakdown is to ground-truth energy breakdown. We calculate the percentage of energy correctly assigned (PEC), where, PEC for the home, appliance, month ($< h, n, m >$) triplet is given by:

$$PEC(h,n,m) = \frac{|E(h,n,m) - \hat{E}(h,n,m)|}{\hat{E}(h, aggregate, m)} \times 100\% \tag{5}$$

where $E_(h,n,m)$ and $\hat{E}(h,n,m)$ denote the predicted and ground-truth usage by appliance n in home h in month m and $\hat{E}(h, aggregate, m)$ denotes the ground truth aggregate home energy usage for home h in month m. The RMS error in the percentage of energy correctly assigned (PEC), for an appliance n is given as the RMS of $PEC(h,n,m)$ across different months and homes,

$$RMS\ PEC(n) = \sqrt{\frac{\sum_h \sum_m PEC(h,n,m)^2}{M \times T}} \tag{6}$$

where M and T indicate the number of homes and months respectively. Lower RMS error in percentage of energy correctly assigned (PEC) means better prediction.

While this metric would allow us to evaluate the performance on a per-appliance basis; we introduce a single metric which is a weighted sum of PECs for different appliances. The weighting is done by the proportion of energy contributed by each appliance in a particular region. Such scheme would require that appliances which contribute more energy to the aggregate are more accurately estimated,

$$Weighted\ PEC = \frac{\sum_n (Frac(n) \times RMS\ PEC(n))}{\sum_n Frac(n)} \tag{7}$$

D. Experimental setup

Our two main experiments involve investigating the performance of transfer learning from source region Austin to target region San Diego homes, and vice-versa. Both these cities have a very different appliance energy breakdown as shown in Figure 1. When transferring **A** from Austin to San Diego, we use all the 534 homes from Austin to learn **A** and testing on all 39 homes from San Diego. However, when we perform the reverse experiment – transferring from San Diego to Austin, we only use 40 test homes from Austin. The rationale is that if we have 500+ homes in a target region, and only around 40 from the source region, it would defeat the purpose of transfer learning. The 40 homes from Austin used for testing were randomly selected, and we repeated this procedure ten times to avoid bias in data sampling.

As done in prior literature [5], we perform our analysis on six appliances – heating, ventilation and air-conditioning (HVAC), fridge, washing machine (WM), microwave (MW), dish washer (DW) and oven. All of these appliances have data from a significant number of homes across both San Diego and Austin. Further, these appliances also represent a wide variety of appliances. For example- season dependent (HVAC) v/s season independent (DW), background (fridge) v/s interactive (WM), etc.

Both TTF and STF involve solving $\text{Min}||E - HAS||$. STF can be solved via implementations of Canonical Polyadic Decomposition (CPD or PARAFAC) [14], [15] or via any of the more recent implementations [16]. However, we found that our implementation, which we used to solve TTF using Autograd for gradient computation, and Adagrad as the optimisation algorithm, provides the best performance over our datasets, and thus we used our projected gradient based method for solving both STF and TTF.

We use nested-cross validation across all our baselines and our approach. For the outer loop (looping across homes), we use 10-fold cross validation. The central point of investigation is: how much adaptation data do we need to beat the baselines? Thus, instead of using 9 folds for training and testing on the 10^{th} fold, we train on $x\%$ data from the 9 folds, where $x\%$ denotes the percentage of adaptation data used. We varied x in {6, 7, 8, 9, 10, 20, 30 ..., 100}. The rationale behind starting with $x = 6\%$ and not a lower number is that we want at least a few homes in the inner loop validation set. We randomly choose the $x\%$ of adaptation homes from the 9 folds for 5 times. This would help reduce the variance in the training set. For the inner loop, we use 2-fold cross-validation. The inner loop is used for parameter/hyper-parameter fine tuning.

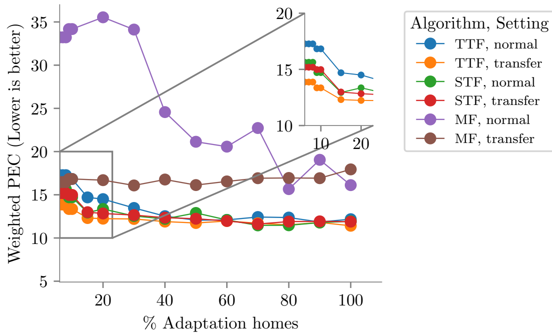

Fig. 2. Our proposed <TTF, Transfer> beats MF and STF baselines for Austin to San Diego transfer.

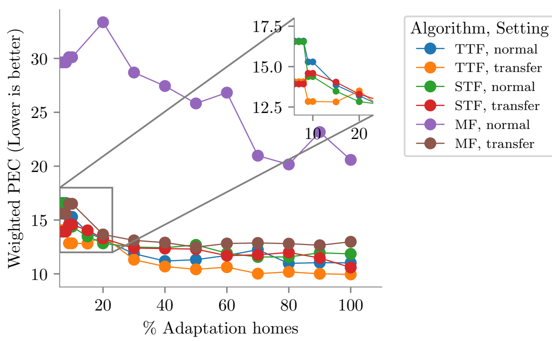

Fig. 3. Our proposed <TTF, Transfer> beats MF and STF baselines for San Diego to Austin transfer.

The set of parameters in TTF (both normal and transfer) are: number of home and season factors; and the hyper-parameters are: the learning rate and the number of iterations. The candidate set of hyper-parameters for STF is the same as that of TTF. For the STF, there is only one parameter – the rank (r).

For the MF based baselines, we used the CVXPy [17] based implementation used by the paper authors. Their implementation solved the MF problem via alternating least squares. The set of parameter for <MF, Transfer> and <MF, Normal> is the number of latent factors. The set of hyper-parameters is the number of iterations of the alternating least squares.

Finally, the $Frac(n)$ required in Eq (7) are used from Table I.

Our entire codebase, baselines, analysis and experiments can be found on Github: https://github.com/nipunbatra/transferable-energy-breakdown.

E. Results and Analysis

Our main result in Figure 2 shows that our approach <TTF, Transfer> performs favourably when compared to all the other baselines on Austin to San Diego transfer. Its most significant advantage over the baselines occurs for low % adaptation homes, which highlights the efficacy of our approach in regions with little instrumentation. Only around 60% adaptation data does the normal learning's error rates approach the error rate of <TTF, Transfer>. This trend is expected since no transfer is perfect, and with sufficient amount of training data available in the target domain, normal learning might do just as well. We can observe that <STF, Normal> does better than <TTF, Normal>. We believe this is due to the trade-off between STF and TTF with respect to the number of parameters. STF requires fewer parameters and is less prone to over-fitting when evaluated in a single region. Both <MF, Transfer> and

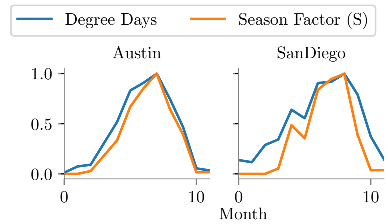

Fig. 4. One of the season factors learnt in our approach corresponds to the amount of air conditioning required

<MF, Normal> show poor accuracy. While <MF, Normal> shows significant improvements as we increase the % of adaptation homes, the error rates remain higher than our proposed approach. It should be mentioned that previously <MF, Normal> had only been tested using a larger dataset [5]. <MF, Transfer> shows little improvement wrt an increase in the % adaptation homes. We believe this is because the learnt appliance-season factor from the source domain has a strong bias towards the seasonal behaviour of the source domain.

On similar lines, results in Figure 3 show that our approach <TTF, Transfer> performs favourably compared to all the other baselines on San Diego to Austin transfer. Unlike the earlier results from Austin to San Diego transfer, in this case, none of the baselines seem to give comparable accuracy beyond 25% adaptation homes.

*1) Intuitive understanding into **S** factors:* We now look at the **S** factors learnt from our approach. In Figure 4, we look at one of the season factors learnt from San Diego and Austin under normal learning setting. We can observe that across both regions, **S** increases in the summer months and has a low value in the winter months. We also plotted the cooling degree days (CDD) in the respective regions and found a high correlation (Pearson coefficient of appx. 0.98) between this season factor and CDD. CDD is a metric to measure the amount of cooling required in a particular region[1]. Without any external supervision, our approach can learn physically relevant parameters such as a factor encoding CDD. The HVAC energy consumption is highly correlated with CDD, and given that we can learn a season factor proportional to CDD, suggests that for HVAC we can learn a region-independent model. Our approach also learns a season factor that is appx, weather independent (i.e. constant across different months). Such a season factor can encode the seasonal energy consumption of appliances such as washing machines, which do not have a direct relationship with CDD.

V. LIMITATIONS

An important limitation of our work is that it will only work well if the target domain has a similar set of appliances as the source domain. If there is an unseen appliance in the target domain, we can not estimate its energy consumption. It must be pointed out that all the baselines discussed also have the same limitation.

VI. CONCLUSIONS

The energy breakdown community has been looking at ways to scale across a large number of homes. One of the major bottlenecks has been that we need data from each region of interest to be able to provide an energy breakdown in that region. We believe that ours is the first approach which does not or rather requires a tiny number of homes from a target region to produce an energy breakdown. Since our evaluation proved that our approach performs favourably compared to the state-of-the-art, we believe that our approach has the potential to help scale energy breakdown.

[1]http://www.degreedays.net/

References

[1] N. Batra, Y. Jia, H. Wang, and K. Whitehouse, "Transferring decomposed tensors for scalable energy breakdown across regions," in *AAAI 2018*, 2018.

[2] S. DeBruin, B. Ghena, Y.-S. Kuo, and P. Dutta, "Powerblade: A low-profile, true-power, plug-through energy meter," in *Sensys 2015*, 2015.

[3] X. Jiang, S. Dawson-Haggerty, P. Dutta, and D. Culler, "Design and implementation of a high-fidelity ac metering network," in *Information Processing in Sensor Networks, 2009. IPSN 2009. International Conference on*. IEEE, 2009, pp. 253–264.

[4] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[5] N. Batra, H. Wang, A. Singh, and K. Whitehouse, "Matrix factorisation for scalable energy breakdown." in *AAAI*, 2017, pp. 4467–4473.

[6] N. Batra, A. Singh, and K. Whitehouse, "Gemello: Creating a detailed energy breakdown from just the monthly electricity bill," in *SIGKDD 2016*, 2016. [Online]. Available: http://doi.acm.org/10.1145/2939672.2939735

[7] ——, "Systems and analytical techniques towards practical energy breakdown for homes," *PhD thesis*, 2017.

[8] R. A. Harshman, "Foundations of the parafac procedure: models and conditions for an" explanatory" multimodal factor analysis," 1970.

[9] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.

[10] O. Parson, G. Fisher, A. Hersey, N. Batra, J. Kelly, A. Singh, W. Knottenbelt, and A. Rogers, "Dataport and nilmtk: A building data set designed for non-intrusive load monitoring," in *GlobalSIP 2015*. IEEE, 2015.

[11] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, no. Jul, pp. 2121–2159, 2011.

[12] D. Maclaurin, D. Duvenaud, and R. P. Adams, "Autograd: Effortless gradients in numpy," in *ICML 2015 AutoML Workshop*, 2015.

[13] C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural computation*, vol. 19, no. 10, pp. 2756–2779, 2007.

[14] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[15] R. Bro, "Parafac. tutorial and applications," *Chemometrics and intelligent laboratory systems*, vol. 38, no. 2, pp. 149–171, 1997.

[16] V. Kuleshov, A. Chaganty, and P. Liang, "Tensor factorization via matrix factorization," in *Artificial Intelligence and Statistics*, 2015, pp. 507–516.

[17] S. Diamond and S. Boyd, "Cvxpy: A python-embedded modeling language for convex optimization," *Journal of Machine Learning Research*, 2016.