Matrix Factorisation for Scalable Energy Breakdown

Nipun Batra[1], **Hongning Wang**[2], **Amarjeet Singh**[1], **Kamin Whitehouse**[2]
nipunb@iiitd.ac.in, hw5x@virginia.edu, amarjeet@iiitd.ac.in, whitehouse@virginia.edu
[1]IIIT Delhi, India, [2] University of Virginia, USA

Abstract

Homes constitute more than one-thirds of the total energy consumption. Producing an *energy breakdown* for a home has been shown to reduce household energy consumption by up to 15%, among other benefits. However, existing approaches to produce an energy breakdown require hardware to be installed in each home and are thus prohibitively expensive. In this paper, we propose a novel application of feature-based matrix factorisation that does not require any additional hardware installation. The basic premise of our approach is that common design and construction patterns for homes create a repeating structure in their energy data. Thus, a sparse basis can be used to represent energy data from a broad range of homes. We evaluate our approach on 516 homes from a publicly available data set and find it to be more effective than five baseline approaches that either require sensing in each home, or a very rigorous survey across a large number of homes coupled with complex modelling. We also present a deployment of our system as a live web application that can potentially provide energy breakdown to millions of homes.

Introduction

Residential buildings are one of the largest energy consumers worldwide, constituting roughly one-thirds of total energy usage (Pérez-Lombard, Ortiz, and Pout 2008). Some of this energy could be saved by producing an *energy breakdown* that itemises the energy consumption of individual loads in the home, such as heating/cooling, lighting, water heating, and refrigeration. An energy breakdown enables informed decision making by several actors in the home's energy ecosystem (Armel et al. 2013). For example, studies show that occupants with access to an energy breakdown can reduce their energy consumption by up to 15% (Kelly and Knottenbelt 2016; Armel et al. 2013). Energy breakdown can also help power utilities and policy makers to improve load forecasting (Armel et al. 2013), to detect broken or misconfigured equipment (Katipamula and Brambley 2005), and to target the most inefficient homes for energy efficiency programs (Armel et al. 2013). Despite the potential benefits, however, only a small number of homes currently have the hardware necessary to create an energy breakdown.

Most homes are not instrumented to produce an energy breakdown because the instrumentation is expensive. A high-frequency smart meter or sub-metering in a home costs up to \$500 per home[1]. The research community has been trying for decades to address the cost of instrumentation through lower-cost sensor designs (DeBruin et al. 2015), data fusion algorithms (Srinivasan, Stankovic, and Whitehouse 2013), and *non-intrusive load monitoring (NILM)*: the use of source separation techniques to estimate the energy consumption of individual loads based on the aggregate power consumption of the entire building (Hart 1992; Armel et al. 2013). However, all of these approaches still require hardware to be installed in every home and therefore have inherent scalability issues. Even if hardware costs were reduced, the cost of labour for installation and maintenance would remain prohibitive. The scalability challenge demands new instrumentation-free approaches.

In this paper, we propose an approach for energy breakdown that does not require any additional hardware installation. The basic premise of our approach is that common design and construction patterns for homes create a repeating structure in their energy data. Thus, a sparse basis in a lower dimensional space can be learned and used to represent energy data from a broad range of homes. A model of a home can be constructed from this basis using only a small amount of easy to collect data, such as utility meter readings, climate zone, and square footage. This low-dimensionality model can then be used to reconstruct sensor data for the home based on high-fidelity data collected in other homes. A recent work called Gemello (Batra, Singh, and Whitehouse 2016) demonstrated the validity of this intuition at small scale.

Our proposed work leverages the advances in the domain of collaborative filtering through feature-based matrix factorisation (Rendle et al. 2011) to the problem of energy breakdown. Since we rely only on monthly bills for energy breakdown, our input for a test home consists of historical monthly bills and some static household properties such as area and the number of occupants. Given that energy is a non-negative quantity, we perform non-negative matrix factorisation on a matrix containing the appliance energy consumption and the aggregate energy consumption across dif-

[1]`http://bit.ly/28UKP62;`

ferent months. We explicitly include the static household properties as known features to guide the factorisation. Including the aggregate energy consumption into the matrix structure helps to address the *cold-start* problem- predicting appliance energy consumption for a home having no previous appliance level data.

We evaluate our approach using 516 homes from the publicly available Dataport data set (Parson et al. 2015), in which the ground truth energy breakdown is measured by metering each appliance of the home individually. Results show that the accuracy of our approach is better or comparable to state-of-the-art NILM technique called latent bayesian melding (LBM), a gold standard NILM technique called factorial hidden Markov model (FHMM), an NILM technique optimised for low-frequency data called discriminative disaggregation sparse coding (DDSC), Gemello and another baseline called regional average. Four of these five baselines (except Gemello) either require sensing in each home, or a very rigorous survey across a large number of homes coupled with complex modelling. We analyse the learnt latent factors and find them to represent relevant physical contexts such as the air conditioning requirement. We also analyse and find that the addition of static household properties helps improve the energy breakdown accuracy.

We used the results from this study to produce an open prototype of the system: a web application[2] that can potentially provide energy breakdown for millions of homes across the US. The web service can take the address of a home and combine static household characteristics from publicly available APIs with the monthly energy bills that can be downloaded through the US Department of Energy's Green Button initiative[3]. This information is combined to estimate an energy breakdown for the household based on sub-metering data from publicly available datasets. As more data becomes publicly available over time, this web service will be able to provide energy breakdowns to more homes and with higher accuracy.

Related work

The research community has tried to improve the scalability of energy breakdowns since George Hart's pioneering work in the 1980s (Hart 1992). These efforts broadly fall into two classes: 1) developing metering hardware that reduces hardware or installation costs, and 2) energy disaggregation techniques such as NILM that estimate the energy of individual loads based on a single aggregate power metering, allowing many loads to be metered with one sensor. However, all existing techniques for sub-metering scale linearly across buildings: at least some hardware must be installed in a new building.

Direct load metering approaches require a power meter for each load. Research prototypes have demonstrated the value of metering hardware that is smaller and cheaper (DeBruin et al. 2015) and easier to install (Donnal and Leeb 2015). Indirect metering approaches use ambient sensors such as light, sound, temperature (Jain, Singh, and Chandan 2016), vibration, and EMI (Gupta, Reynolds, and Patel 2010; Gulati et al.), and to help infer the activity of a load.

Disaggregation is a general class of techniques that infer the power or energy of individual loads based on aggregate power measurements of the entire building. Disaggregation is an under-constrained problem and therefore requires source separation techniques that use some model of the energy loads (Kolter and Jaakkola 2012; Parson et al. 2012; Hart 1992; Zhong, Goddard, and Sutton 2015). Many algorithms require training data for each load (Hart 1992; Kolter and Jaakkola 2012) while others learn the model in an unsupervised fashion (Barker et al. 2013; Shao, Marwah, and Ramakrishnan 2013). Disaggregation techniques sacrifice metering accuracy for the advantage of having only a single meter per building. However, this approach still has the scalability problem because the metering hardware required is currently installed in very few buildings. Most disaggregation techniques require aggregate power metering at 1/60Hz or greater, and sometimes at frequencies of 10KHz or higher (Gupta, Reynolds, and Patel 2010; Berges et al. 2010). However, most meters installed today provide metering data only once per month. Advanced metering infrastructure (AMI) is currently installed in about 30% of households worldwide[4], but mostly provides metering data at a 15-minute or higher sampling interval. Current algorithms generally give poor disaggregation accuracies at such low frequencies. Thus, the specialised hardware required for disaggregation still scales linearly with the number of buildings and cannot leverage smart power meters that are already installed.

Prior work (Batra, Singh, and Whitehouse 2016) demonstrated the validity of the intuition of our work. Their system called Gemello estimated the energy usage of one home based on other homes that were very similar and used kNN type algorithms for matching. However, advances in the problem of collaborative filtering, most notably the Netflix prize (Koren et al. 2009) have shown the superiority of matrix factorisation methods over kNN type methods. Our work is inspired by recent advances in the collaborative filtering domain. While there exists some work that applies similar techniques for energy breakdown (Kolter, Batra, and Ng 2010; Wytock and Kolter 2014), these works are applicable on data coming from a smart meter and thus requires some hardware in the test home.

It must be clarified that both our approach and Gemello can produce an energy breakdown only at a monthly temporal resolution, unlike NILM solutions that can produce a high frequency appliance energy time-series. However, previous studies have shown sustained savings even at a monthly resolution (Kelly and Knottenbelt 2016).

Approach- Matrix Factorisation (MF)

The overall goal of our matrix factorisation (MF) based approach is to predict per-appliance energy consumption in a test home, without requiring any sensing instrumentation, given the per-appliance energy consumption across some

[2]https://github.com/nipunbatra/scalable-energy-breakdown-webapp

[3]http://www.greenbuttondata.org/

[4]http://bit.ly/2cngqxf

small number of training homes. The basic premise of our approach is that common design and construction patterns for homes create a repeating structure in their energy data. Thus, a sparse basis can be learned and used to represent energy data from a broad range of homes. A model of a home can be constructed from this basis using only a small amount of data, such as utility meter readings, climate zone, and square footage. This low-dimensionality model can then be used to reconstruct sensor data for the home based on high-fidelity data collected in other homes.

For each appliance $\mathbf{i}$, we create a matrix $\mathbf{X_i} \in \mathbf{R}^{\mathbf{m}\times\mathbf{2n}}$, where $\mathbf{m}$ corresponds to different homes, and there are $\mathbf{2n}$ columns- $\mathbf{n}$ coming from home aggregate energy over different months and $\mathbf{n}$ coming from appliance energy over different months. Our goal is to predict the per-appliance energy consumption of a home while observing only the aggregate monthly bill for the home, alongside some static properties, such as area and number of occupants. For a test home, the $\mathbf{n}$ entries in $\mathbf{X_i}$ corresponding to appliance energy across months will be absent (and need to be predicted). The $\mathbf{n}$ entries in $\mathbf{X_i}$ from household aggregate energy across different months help to predict appliance energy for this home. We now discuss several properties and insights in designing matrices and solving MF for our problem:

1. Non-negative constraints: Energy is a non-negative quantity. Thus, this formulation should be posed as non-negative matrix factorisation (NNMF) (Lee and Seung 2001). Thus, for the $\mathbf{i^{th}}$ appliance, when using $\mathbf{k}$ latent factors, we aim to learn $\mathbf{A_i} \in \mathbf{R}^{\mathbf{m}\times\mathbf{k}}$ and $\mathbf{B_i} \in \mathbf{R}^{\mathbf{k}\times\mathbf{2n}}$, such that $\mathbf{X_i} \approx \mathbf{A_i}\mathbf{B_i}$, where $\mathbf{A_i} \geq \mathbf{0}$, $\mathbf{B_i} \geq \mathbf{0}$ and $\mathbf{k} < \mathbf{m}, \mathbf{2n}$. This can be formulated as an optimisation problem:

$$\begin{aligned} \text{Min } ||\mathbf{X_i} - \mathbf{A_i}\mathbf{B_i}||_\mathbf{F}^\mathbf{2} + \lambda_\mathbf{1}||\mathbf{A_i}||_\mathbf{2}^\mathbf{2} + \lambda_\mathbf{2}||\mathbf{B_i}||_\mathbf{2}^\mathbf{2} \\ \text{s.t. } \mathbf{A_i}, \mathbf{B_i} \geq \mathbf{0} \end{aligned} \quad (1)$$

where $\lambda_\mathbf{1}, \lambda_\mathbf{2}$ are regularisation parameters, $||\mathbf{Y}||_\mathbf{F}$ indicates the Frobenius norm and $||\mathbf{y}||_\mathbf{2}$ indicates the $\mathbf{l_2}$ norm. $\mathbf{A_i}$ corresponds to latent factor for homes and may relate to properties of a home impacting energy usage, such as insulation level, area of the home, among others. $\mathbf{B_i}$ corresponds to the latent factor for months and may relate to energy consumption of the $\mathbf{i^{th}}$ appliance as a function of seasons.

2. Incorporating household features: Static features such as area of home, number of occupants are often correlated with appliance usage, and if known can be explicitly specified as known factors to guide the factorisation. Prior literature has shown that such feature-based factorisation is more accurate than conventional latent factor models (Rendle et al. 2011). Thus, given a matrix $\mathbf{D} \in \mathbf{R}^{\mathbf{m}\times\mathbf{d}}$ containing data for $\mathbf{d}$ static household properties, we modify our factorisation model from $\mathbf{X_i} \approx \mathbf{A_i}\mathbf{B_i}$ to $\mathbf{X_i} \approx \mathbf{A_i}\mathbf{B_i} + \mathbf{D}\theta^\mathbf{T}$, where θ is the shared regression coefficient across homes.

Our final formulation for the i^{th} appliance can be written as:

$$\begin{aligned} \text{Min } ||\mathbf{X_i} - (\mathbf{A_i}\mathbf{B_i} + \mathbf{D}\theta^\mathbf{T})||_\mathbf{F}^\mathbf{2} + \lambda_\mathbf{1}||\mathbf{A_i}||_\mathbf{2}^\mathbf{2} + \lambda_\mathbf{2}||\mathbf{B_i}||_\mathbf{2}^\mathbf{2} \\ \text{s.t. } \mathbf{A_i}, \mathbf{B_i} \geq \mathbf{0} \end{aligned} \quad (2)$$

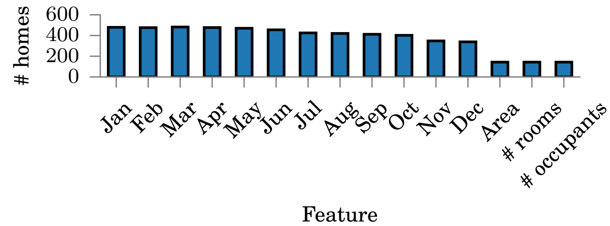

Figure 1: Variable number of features are available across 516 homes in our data set.

At this point, we would like to clarify that a matrix structure where all appliances are considered (Kolter, Batra, and Ng 2010), i.e. a matrix of the shape $\mathbf{m} \times (\mathbf{L} \times \mathbf{n})$, where $\mathbf{L}$ is the number of considered appliances, may or may not result in better disaggregation. This is due to the fact that the latent factor for homes may not be shared across appliances. Testing on our data set revealed that our matrix structure of $\mathbf{m} \times \mathbf{2n}$ gives better or comparable performance to the matrix structure of $\mathbf{m} \times (\mathbf{L} \times \mathbf{2n})$, while being quicker to factorise. We defer a detailed analysis of the trade-off between these two matrix structures for future work.

Evaluation

Dataset

We use the publicly available Dataport (Parson et al. 2015) data set for evaluation. Dataport is the largest[5] public data set for household energy data. Dataport data set has data from 586 homes in Austin, Texas, USA for the year 2015. Power data is logged every minute for household aggregate and multiple appliances in this data set. The data set also contains static household properties such as household area, number of occupants, and number of rooms for a subset of the homes. We filter out 70 homes that do not have aggregate energy consumption for even a single month. Of the remaining 516 homes, 105 homes have all available features (12 month household aggregate energy and 3 static features- area, number of occupants, number of rooms). Figure 1 shows the distribution of features across homes.

Baselines

We compare the accuracy of our approach against the following five baselines.

Regional average (RA): The US Energy Information Administration (EIA) conducts the residential energy consumption survey (RECS) every 5 years. They use a fairly involved process to estimate the contribution of different appliances to energy consumption across different regions. This includes surveys across tens of thousands of homes to capture energy characteristics, followed by building non-linear statistical models from household monthly energy bills to estimate the energy consumption across different appliances. For RA baseline, we compute the predicted energy usage of an appliance in a region as the product of the regional average proportion of that appliance and the aggregate monthly energy consumption of the home.

[5] http://bit.ly/28Xnlju

NILM- FHMM, LBM and DDSC: We use three NILM techniques as baselines. We use a factorial hidden Markov model (FHMM) (Ghahramani and Jordan 1997; Kolter and Jaakkola 2012), which is accepted as a gold standard in NILM literature. In an FHMM, each appliance is modelled as a Gaussian hidden Markov model, containing three parameters: prior, transition matrix and emission matrix. Each appliance is modelled to contain S states (such as ON, OFF, etc.). The prior encodes the initial probability of an appliance starting in different states ($\{1..S\}$). The transition matrix encodes the probability of transition from state s_i to s_j. The emission matrix encodes the distribution of power for different states.

We use the state-of-the-art NILM technique based on latent Bayesian melding (LBM) (Zhong, Goddard, and Sutton 2015; 2014), as our second NILM benchmark. The goal of this work by Zhong et. al is to break down the energy consumption into appliances given the aggregate power time series . The underlying model used in this approach is an FHMM. In addition to modelling the system as an FHMM, the authors in this work add prior constraints to improve the accuracy. An example of such constraints is the expected number of ON/OFF transitions of an appliance. We use discriminative disaggregation sparse coding (DDSC) (Kolter, Batra, and Ng 2010) as the third NILM baseline. DDSC is based upon structured prediction for discriminatively training sparse coding algorithms specifically to maximise disaggregation performance.

All these three NILM technique produce a high frequency time series for different appliances and we sum up the energy consumption to obtain per-appliance monthly energy consumption.

Gemello/kNN We use Gemello (Batra, Singh, and Whitehouse 2016) as our final baseline. Gemello in its direct form is applicable only to homes having all features and thus we can only apply this baseline to the subset of homes satisfying this constraint. For the remaining homes, having a variable number of features, we use kNN where distances between homes are calculated based on common set of features. It must be pointed that we could have alternatively imputed the missing entries and used Gemello. We keep such an analysis for the future.

Implementation of our approach

The optimisation proposed for our approach proposed in Equation 2 is not jointly convex in $\mathbf{A_i}$ and $\mathbf{B_i}$. However, by fixing one, the optimisation becomes convex in the other, which we solve using an alternating least square (ALS) strategy. Another important implementation detail involves linearly normalising the matrix entries on a scale of 0 to 1 by using the maximum and the minimum entry in the matrix.

Evaluation metric

We chose our metric after deliberating on the metrics used in prior work and our discussions with NILM experts. Since different appliances are on a different scale (HVAC consumes significantly more energy than a microwave), comparing the RMS error in energy consumption can be hard to interpret across appliances. Normalising the error by actual usage may seem a possible solution. However, this metric breaks for low-energy appliances. For example, if the actual and predicted usage of the oven is 0.1 and 0.2 units, error would be 100%. However, an error of 0.1 units would probably be insignificant in absolute terms. To overcome the problems of the above two metrics, we choose a metric defined as RMS error in percentage of energy correctly assigned (PEC) (Batra et al. 2014), where, PEC for the home, appliance, month ($< h, w, m >$) triplet is given by:

$$PEC(h, w, m) = \frac{|w_{prediction}(h, m) - w(h, m)|}{aggregate(h, m)} \times 100\% \quad (3)$$

where $w(h, m)$ denotes the ground truth energy usage by appliance w in home h in month m and $aggregate(h, m)$ 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 w is given as the RMS of $PEC(h, w, m)$ across different months and homes. Lower RMS error in percentage of energy correctly assigned (PEC) means better prediction.

HVAC	Fridge	Washing machine	Dishwasher
0.29	0.09	0.01	0.02

Table 1: Proportion of energy consumed by different appliances in Austin.

Experimental setup

We perform our analysis on six appliances - heating, ventilation and air-conditioning (HVAC), fridge, washing machine (WM), microwave (MW), dish washer (DW) and oven. There are three main reasons for choosing these six appliances. First, our data set contains a substantial number of homes with these 6 appliances. Second, these six appliances represent a diverse category: i) HVAC represents appliances that are heavily affected by weather and consume high energy, ii) fridge represents always ON appliances, that are moderately affected by weather and usage, iii) washing machine and dryer represents appliances that are highly usage dependent and typically consume low energy relative to HVAC and fridge, oven and microwave represent appliances used in the kitchen. Third, together these six appliances contribute more than half of the total household energy. We perform our evaluation on two different test sets- 105 homes having all feature and 516 homes containing homes with missing features.

For regional average (RA) baseline, we use the numbers obtained from RECS survey as shown in Table 1. It must be noted that the RECS survey does not have appliance level numbers for oven and microwave, and we thus can not make a prediction for these two appliances using RA baseline.

For our FHMM and LBM baselines, we use their implementation in NILMTK (Batra et al. 2014; Kelly et al. 2014) and model each appliance as a 3-state appliance (Off, Intermediate and High power), as per the work in (Zhong, Goddard, and Sutton 2015). To measure the NILM performance given current smart meters, we feed the NILM algorithm 15-minute aggregate reading which it tries to break down

	FHMM	LBM	DDSC	RA	Gemello	MF
HVAC	15.26	29.37	31.39	17.44	12.62	**12.53**
Fridge	4.48	**2.69**	4.32	4.62	4.37	3.65
Oven	34.09	3.84	1.37	-	1.07	**1.04**
DW	12.99	1.74	1.30	1.22	1.05	**0.92**
WM	3.98	13.29	1.36	0.71	0.50	**0.49**
MW	6.32	1.01	1.08	-	0.87	**0.64**

Table 2: RMS error (lower is better) in the percentage of energy assigned for 105 homes having all features.

	FHMM	LBM	RA	KNN	MF
HVAC	15.65	29.37	18.40	**11.96**	12.02
Fridge	3.90	**2.69**	4.41	3.38	3.62
Oven	34.00	3.84	-	1.49	**1.32**
DW	13.80	1.74	1.22	1.01	**0.92**
WM	3.89	13.29	1.40	1.45	**1.33**
MW	5.76	1.01	-	0.98	**0.91**

Table 3: RMS error (lower is better) in the percentage of energy assigned for 516 homes (having missing features).

into 15-minute time series for the six appliances. The NILM model is trained on the entire 516 homes including the test homes as we wanted to see the best performance of baseline algorithms. Due to time constraints, we were able to evaluate the performance of DDSC only over the 105 homes having all features. DDSC was inputted 15-minute appliance and aggregate power traces for training and 15-minute home aggregate power traces for testing. Optimal parameters for DDSC were learnt using cross-validation. The three NILM approaches produce as output a 15-minute power time series for each appliance which is aggregated to monthly appliance energy consumption.

Gemello has top-N features and number of neighbours K as tunable parameters. For Gemello, we use the parameters used in previous work (Batra, Singh, and Whitehouse 2016), K varies from 1 to 6, and N varies from 1 to 8.

Our MF based approach has regularisation (λ), static features to include (area, number of occupants and number of rooms) and the number of latent factors as the tunable parameters. We varied λ in factors of 10 from 10^{-3} to 10^{2}. We used all length-0, 1, 2 and 3 combinations of the 3 static features (<None>, <area>, <#occupants>,… <area, #occupants, #rooms>). We varied the number of latent factors from 1 to 10.

For both Gemello and MF, we use a nested *leave-one-out* cross-validation strategy. The inner loop is used to fine-tune the parameters. The outer loop is used for prediction of energy across different appliances for a test home, when all but that home are used in the train set. It must be pointed out that both Gemello and our MF approach have the same set of input information available (historical aggregate energy and appliance montly energy consumption, and three static household properties). Our entire implementation, experiments and analysis can be found on Github (URL not mentioned for anonymity).

Results and Analysis

Our main result in Table 2 on 105 homes having all features, shows that our MF approach gives better energy breakdown performance than the four baselines for 5/6 appliances. The relative improvement in energy breakdown performance over the best baseline, is the highest for microwave and dish washer. Both these appliances are generally considered problematic for traditional NILM algorithms (Barker et al. 2013) owing to their multiple states of operation and in general sparse usage. For the fridge, LBM gives best performance followed by our approach. This may be due to the fact that LBM is accurately able to balance the prior (expected number of cycle and energy usage) with the time series data for the fridge. Other appliances may not be showing such cyclic behaviour.

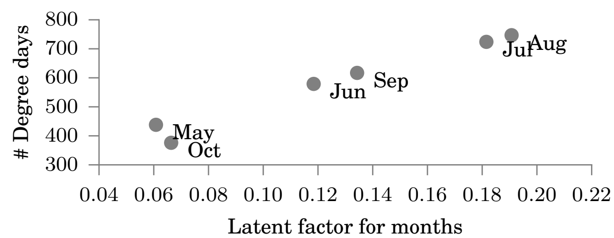

Figure 2: One of the latent factors learnt for HVAC has a high correlation with the # of degree days

In Table 3, we see that our MF approach gives better energy breakdown performance than the four baselines for 4/6 appliances for 516 homes. As we saw before, LBM does best for the fridge. For HVAC, while KNN gives the best performance, our approach is comparable.

We now analyse the efficacy of our MF based approach on the data from 105 homes. When learning latent factors for HVAC, we found one of the factors for month to be highly correlated with the air conditioning requirement for that month (Figure 2). The air conditioning requirement for a month can be captured by a parameter called the number of degree days[6]. Since the HVAC energy consumption is seasonal and depends on the number of degree days, our approach is expected to work better than baselines (including KNN), which aren't able to capture such information. On a similar front, when we did MF without explicitly incorporating static features, we found that some of the latent factors had a high correlation with these static parameters. Figure 3 shows the relative gain in performance by the addition of these static features over the standard MF. While all appliances show an improvement in performance by the addition of static features, dish washer has the maximum gain. This is consistent with previous similar work (Batra, Singh, and

[6] https://en.wikipedia.org/wiki/Degree_day

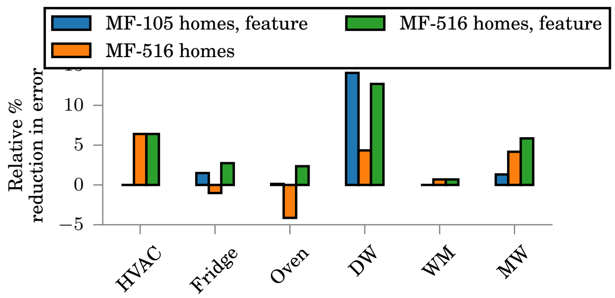

Figure 3: Reduction in error over MF on 105 homes over 6 appliances. Incorporating static features into our matrix factorisation improves energy breakdown performance.

Figure 4: Screenshot from the web user interface that can potentially provide energy breakdown to millions of homes in the US leveraging our approach.

Whitehouse 2016), which shows that static features are useful for appliances such as dish washer.

We further tried to answer the question- *"What's better? More, but incomplete data, or, less but complete data"*. For this, we use all the 516 homes for training and analysed the performance of the test 105 homes having all features, compared to training only on these 105 homes. Our results in Figure 3 show that for 4/6 appliances, the performance improves by adding more homes and performing plain MF (without additional features). When static features are also considered, there is an improvement in performance for all the 6 appliances. While this data may not be sufficient for conclusively saying that more data is better, the case for the value of static features is more conclusive.

Implementation For Scale

We now discuss an implementation of our system which can scale to millions of homes across the US. The US Energy department runs a program called Green Button, under which, more than 50 utilities across the US are allowing 60 million households to download their energy consumption in a standard format. This program caters to users having smart meters and traditional electricity meters. We have created a web application where users can upload their Green Button data to obtain their per-appliance energy breakdown, which we obtain by applying our approach on existing data sets having appliance level data. To obtain household static properties, we request the users for their address and can pull information such as household area and age from online APIs such as the one offered by Zillow[7]. Figure 4 shows a screenshot from an initial prototype.

Discussion

We now discuss two additional properties and insights that can be incorporated into our approach that we did not consider due to space and time constraints. We believe that such domain insights can be captured in the MF formulation.

1. Temporal characteristics: We can categorise household appliances into those affected (e.g. HVAC) or not affected (e.g. oven) by seasonal trends. For appliances not affected by seasonal changes, we can impose a penalty on variation in predicted energy consumption across months. For appliances that are affected by seasonal variations, we can explicitly add properties capturing seasonal variations (such as temperature) as known latent factors for $\mathbf{B}$ (Wytock and Kolter 2014).

2. Appliance correlations: The energy usage of different appliances is often correlated (Kim et al. 2011). For example, the energy usage of a dryer is likely to be correlated with the washing machine. This property can be captured by constructing a matrix structure containing all the correlated appliances as well as aggregate energy. The latent factors can be constrained in a similar fashion as we did in Equation **??**.

Limitations and Future Work

While we have shown that our system provides better energy breakdown in comparison to NILM, there are many applications of NILM which our system cannot provide, enabled by higher frequency of output. Our current work also assumes that data comes from the same geographical location. For homes spanning multiple geographical locations, we need to incorporate knowledge transfer that accounts for the differences between these sets of homes (such as weather). In the future, we plan to extend our approach to incorporate some of the above discussed constraints, such as the correlations that exist between appliances.

Conclusions

Energy breakdown literature has largely looked at methods that require additional hardware to be installed. Due to prohibitive cost, it is unlikely that a significant proportion of the world will have access to such hardware. We believe that our approach presents an interesting dimension to the well-studied problem and owing to the no additional hardware nature, is likely to be easier to scale. All the infrastructure required to scale such an approach already exists. The efficacy of our approach is shown by its competitiveness against state-of-the-art NILM methods that rely on additional hardware.

[7] `http://bit.ly/1PWZGOp`

References

Armel, K. C.; Gupta, A.; Shrimali, G.; and Albert, A. 2013. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy* 52:213–234.

Barker, S.; Kalra, S.; Irwin, D.; and Shenoy, P. 2013. Empirical characterization and modeling of electrical loads in smart homes. In *IEEE IGCC*.

Batra, N.; Kelly, J.; Parson, O.; Dutta, H.; Knottenbelt, W.; Rogers, A.; Singh, A.; and Srivastava, M. 2014. NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring. In *Fifth International Conference on Future Energy Systems*.

Batra, N.; Singh, A.; and Whitehouse, K. 2016. Gemello: Creating a detailed energy breakdown from just the monthly electricity bill. In *SIGKDD 2016*.

Berges, M. E.; Goldman, E.; Matthews, H. S.; and Soibelman, L. 2010. Enhancing electricity audits in residential buildings with nonintrusive load monitoring. *Journal of industrial ecology* 14(5):844–858.

DeBruin, S.; Ghena, B.; Kuo, Y.-S.; and Dutta, P. 2015. Powerblade: A low-profile, true-power, plug-through energy meter. In *Sensys 2015*.

Donnal, J. S., and Leeb, S. B. 2015. Noncontact power meter. *IEEE Sensors Journal* 15(2):1161–1169.

Ghahramani, Z., and Jordan, M. I. 1997. Factorial hidden markov models. *Machine learning* 29(2-3).

Gulati, M.; Singh, V.; Agarwal, S.; and Bohara, V. Appliance activity recognition using radio frequency interference emissions.

Gupta, S.; Reynolds, M. S.; and Patel, S. N. 2010. Electrisense: single-point sensing using emi for electrical event detection and classification in the home. In *Ubicomp*.

Hart, G. W. 1992. Nonintrusive appliance load monitoring. *Proceedings of the IEEE* 80(12):1870–1891.

Jain, M.; Singh, A.; and Chandan, V. 2016. Non-intrusive estimation and prediction of residential ac energy consumption. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 1–9. IEEE.

Katipamula, S., and Brambley, M. R. 2005. Review article: methods for fault detection, diagnostics, and prognostics for building systemsa review. *HVAC Research* 3–25.

Kelly, J., and Knottenbelt, W. 2016. Does disaggregated electricity feedback reduce domestic electricity consumption? a systematic review of the literature. *arXiv preprint arXiv:1605.00962*.

Kelly, J.; Batra, N.; Parson, O.; Dutta, H.; Knottenbelt, W.; Rogers, A.; Singh, A.; and Srivastava, M. 2014. Nilmtk v0. 2: a non-intrusive load monitoring toolkit for large scale data sets: demo abstract. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, 182–183. ACM.

Kim, H.; Marwah, M.; Arlitt, M. F.; Lyon, G.; and Han, J. 2011. Unsupervised Disaggregation of Low Frequency Power Measurements. In *Proceedings of 11th SIAM International Conference on Data Mining*, 747–758.

Kolter, J. Z., and Jaakkola, T. 2012. Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*.

Kolter, J. Z.; Batra, S.; and Ng, A. Y. 2010. Energy Disaggregation via Discriminative Sparse Coding. In *NIPS 2010*.

Koren, Y.; Bell, R.; Volinsky, C.; et al. 2009. Matrix factorization techniques for recommender systems.

Lee, D. D., and Seung, H. S. 2001. Algorithms for non-negative matrix factorization. In *NIPS 2001*.

Parson, O.; Ghosh, S.; Weal, M.; and Rogers, A. 2012. Non-intrusive load monitoring using prior models of general appliance types. In *AAAI 2012*.

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

Pérez-Lombard, L.; Ortiz, J.; and Pout, C. 2008. A review on buildings energy consumption information. *Energy and buildings* 40(3):394–398.

Rendle, S.; Gantner, Z.; Freudenthaler, C.; and Schmidt-Thieme, L. 2011. Fast context-aware recommendations with factorization machines. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, 635–644. ACM.

Shao, H.; Marwah, M.; and Ramakrishnan, N. 2013. A temporal motif mining approach to unsupervised energy disaggregation: Applications to residential and commercial buildings. In *AAAI*.

Srinivasan, V.; Stankovic, J.; and Whitehouse, K. 2013. Fixturefinder: discovering the existence of electrical and water fixtures. In *IPSN*.

Wytock, M., and Kolter, J. Z. 2014. Contextually supervised source separation with application to energy disaggregation. In *AAAI 2014*. AAAI Press.

Zhong, M.; Goddard, N.; and Sutton, C. 2014. Signal aggregate constraints in additive factorial hmms, with application to energy disaggregation. In *NIPS 2014*.

Zhong, M.; Goddard, N.; and Sutton, C. 2015. Latent bayesian melding for integrating individual and population models. In *NIPS 2015*.