Chapter 5

Scalable energy disaggregation

5.1 Introduction

Only a small number of homes have the necessary infrastructure or hardware to support a good amount of work in the academic NILM community. 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 [31], data fusion algorithms [97], 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 [50, 6]. 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 chapter, 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 can be learned and used to represent energy

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

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.

Our work leveraged the advances in the domain of collaborative filtering through feature-based matrix factorisation to the problem of energy breakdown [90]. Since we rely only on monthly bills for energy breakdown, our input 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 different 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 [85], 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 techniques. These baselines either require sensing in each home, or a very rigorous survey across a large number of homes coupled with complex modelling. We analysed the learnt latent factors and found them to represent relevant physical contexts such as the air conditioning requirement. We also analysed and found that the addition of static household properties helps improve the energy breakdown performance.

We used the results from this study to produce an open prototype of the system: a web application that can potentially provide energy breakdown for millions of homes across the US. The web service takes the address of a home and can 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

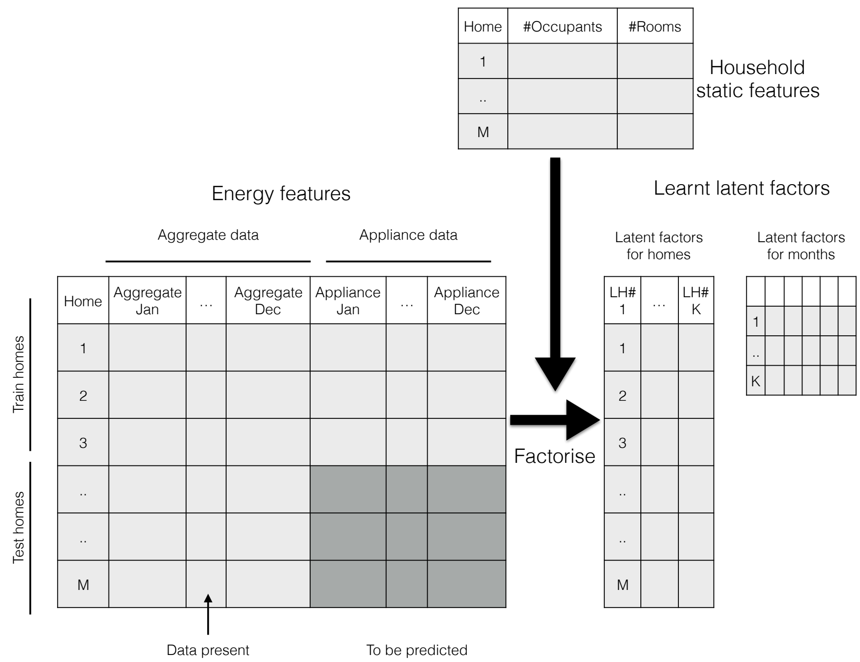

Figure 5-1

initiative[2]. 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.

5.2 Approach- Matrix Factorisation (MF)

The overall goal of our matrix factorisation (MF) (Figure 5-1) 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 small number of train 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

[2] `http://www.greenbuttondata.org/`

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 $\mathbf{X_i} \in \mathbf{R^{m\times 2n}}$, 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 helps to solve the issue of cold-start and 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) [76]. Thus, for the $\mathbf{i^{th}}$ appliance, when using $\mathbf{k}$ latent factors, we aim to learn $\mathbf{A} \in \mathbf{R^{m\times k}}$ and $\mathbf{B} \in \mathbf{R^{k\times 2n}}$, such that $\mathbf{X_i} \approx \mathbf{AB}$, where $\mathbf{A} \geq \mathbf{0}$, $\mathbf{B} \geq \mathbf{0}$ and $\mathbf{k} < \mathbf{m}, \mathbf{2n}$. This can be formulated as an optimisation problem:

$$\text{Min } ||X_i - AB||_F^2 + \lambda_1||A||_2^2 + \lambda_2||B||_2^2 \;\text{ s.t. } A, B \geq 0 \tag{5.1}$$

where λ_1, λ_2 are regularisation parameters, $||\mathbf{Y}||_\mathbf{F}$ indicates the Frobenius norm and $||\mathbf{y}||_\mathbf{2}$ indicates the $\mathbf{l_2}$ norm. $\mathbf{A}$ 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. $\mathbf{B}$ corresponds to the latent factor for months and may relate to energy consumption of an 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 [90]. Thus, given a matrix $\mathbf{D} \in \mathbf{R^{m\times d}}$ containing data for $\mathbf{d}$

static household properties, we modify our factorisation model from $X_i \approx AB$ to $X_i \approx AB + D\theta^T$, where θ is the shared regression coefficient across homes.
Our final formulation for the i^{th} appliance can be written as:

$$\text{Min } ||X_i - (AB + D\theta^T)||_F^2 + \lambda_1||A||_2^2 + \lambda_2||B||_2^2 \text{s.t. } A, B \geq 0 \tag{5.2}$$

At this point, we would like to clarify that a matrix structure where all appliances are considered [72], i.e. a matrix of the shape $\mathbf{m} \times (\mathbf{I} \times \mathbf{n})$, where $\mathbf{I}$ is the number of considered appliances, may or may not result in better disaggregation. This is due to the fact that not all homes may have all appliances and thus for uncommon appliances, the corresponding matrix entries will be mostly sparse. Thus, there is a trade-off between the additional sparseness that negatively affects matrix factorisation and the additional appliance information that may be available for a home, that would likely aid matrix factorisation. 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 $\mathbf{m} \times (\mathbf{I} \times \mathbf{2n})$, while being quicker to factorise. We defer a detailed analysis of the trade-off between these two matrix structures for future work.

Our approach can currently only make accurate predictions for homes in a particular region. In other words, the train and the test homes should come from the same region. The energy patterns across different regions can vary substantially. Thus, if the train data and test data come from different regions, our approach may give poor energy breakdown accuracy. In the future, we plan to address this limitation by transferring knowledge across regions [100].

5.3 Evaluation

5.3.1 Dataset

We use the publicly available Dataport [85] data set for evaluation. Dataport is the largest[3] public data set for household energy data. Dataport data set has data from

[3] `http://bit.ly/28Xnlju`

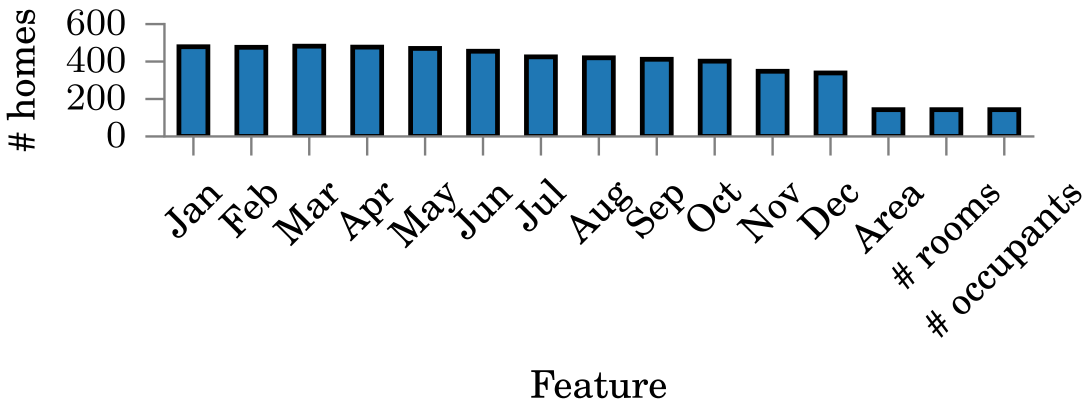

Figure 5-2: Variable number of features are available across 516 homes in our data set.

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 don't 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 5-2 shows the distribution of features across homes.

5.3.2 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.

NILM- FHMM, LBM and DDSC:

We use three NILM techniques as baselines. We use a factorial hidden Markov model (FHMM) [41, 73], 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) [107, 105] proposed by Zhong et. al, 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) [72] 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 [20] as our final baseline. Gemello in its direct form is applicable only to homes having all features and thus we can 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.

5.3.3 Implementation of our approach

The optimisation proposed for our approach proposed in Equation 5.2 is not jointly convex in **A** and **B**. However, by fixing one, the optimisation becomes convex in the other. Thus, we implement an alternating least square (ALS) strategy implemented in Python using CVXPY [33]. CVXPY also allows us to specify the non-negative constraints and incorporating static features. 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.

5.3.4 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) [16], where, PEC for the home, appliance, month ($< h, w, m >$) triplet is given by:

$$PEC(h, w, m) = \frac{|w_{prediction}(h, m) - w(h, m)|}{aggregate(h, m)} \times 100\% \tag{5.3}$$

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

HVAC	Fridge	Washing machine	Dishwasher
0.29	0.09	0.01	0.02

Table 5.1: Proportion of energy consumed by different appliances in Austin.

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.

5.3.5 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 5.1. It must be noted that the RECS survey doesn't have appliance level numbers for oven and microwave, and we thus can't make a prediction for these two appliances using RA baseline.

For our FHMM and LBM baselines, we use their implementation in NILMTK [16] and model each appliance as a 3-state appliance (Off, Intermediate and High power), as per the work in [107]. To measure the NILM performance given current smart meters, we feed the NILM algorithm 15-minute aggregate reading which it tries to break down 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. It must be mentioned that while the LBM implementation comes from the authors of that paper, the FHMM one comes from a publicly available toolkit, the implementation of DDSC is ours and thus may not fully match with the authors' version.

Gemello has top-N features and number of neighbours K as tunable parameters. For Gemello, we use the parameters used in previous work [20], 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. We chose to set 10 as the upper limit on the number of latent factors as we have data from 12 months, and we would want a low-rank approximation.

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).

	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 5.2: 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 5.3: RMS error (lower is better) in the percentage of energy assigned for 516 homes (having missing features).

5.3.6 Results and Analysis

Our main result in Table 5.2 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 [7] 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.

In Table 5.3, 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

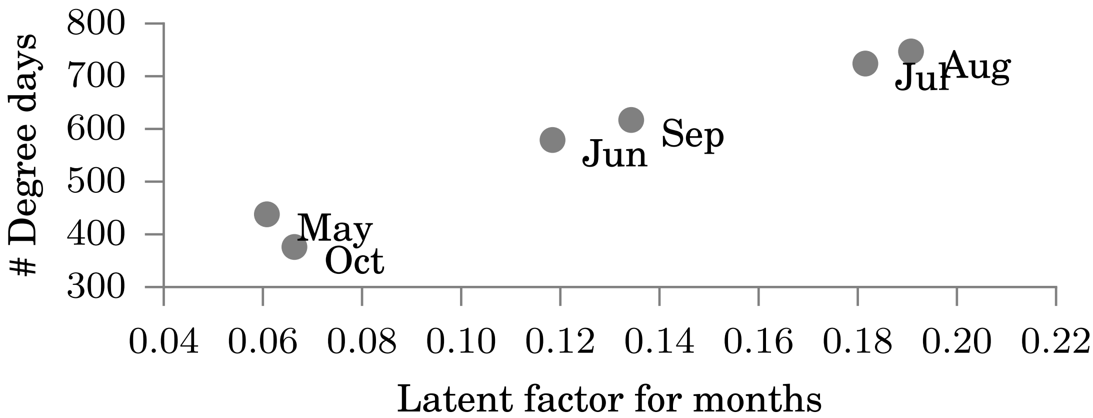

Figure 5-3: One of the latent factors learnt for HVAC has a high correlation with the # of degree days

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 5-3). The air conditioning requirement for a month can be captured by a parameter called the number of degree days[4]. 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 5-4 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 [20], 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 5-4 show that for 4/6 appliances, the

[4] `https://en.wikipedia.org/wiki/Degree_day`

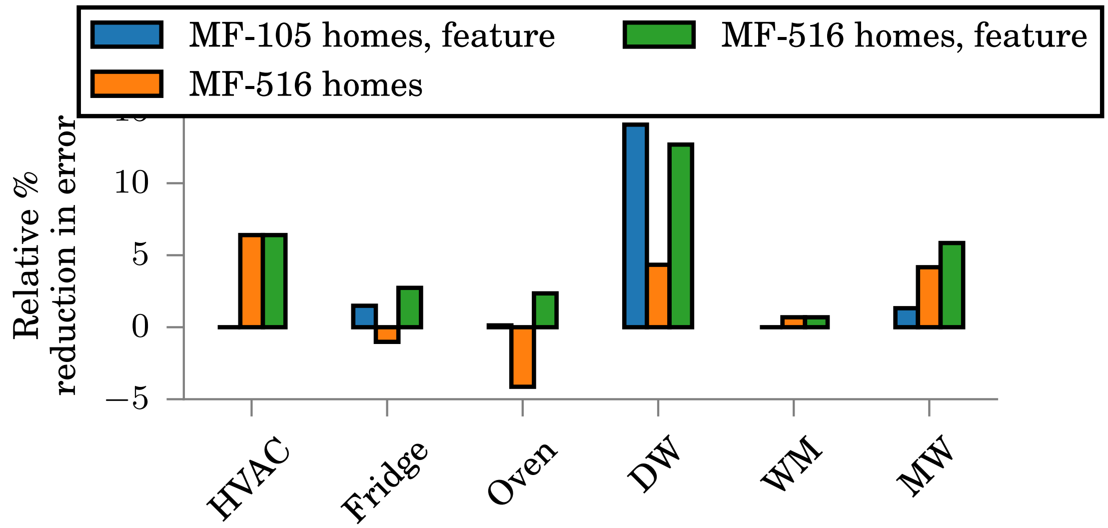

Figure 5-4: Reduction in error over MF on 105 homes over 6 appliances. Incorporating static features into our matrix factorisation improves energy breakdown performance.

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.

5.4 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

Figure 5-5: Screenshot from the web user interface that can potentially provide energy breakdown to millions of homes in the US leveraging our approach.

and age from online APIs such as the one offered by Zillow[5]. Figure 5-5 shows a screenshot from an initial prototype.

5.5 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. Previous work has shown that energy breakdown performance can be improved by incorporating correlation of appliances with seasonal weather data[102] and the correlation between appliances [68]. 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. The penalty can be imposed by adding the

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

following term to Equation 5.1:

$$\text{Min}\ ^{-}\sum_{\mathbf{i=1}}^{\mathbf{k}}\sum_{\mathbf{j=1}}^{\mathbf{2n-1}}(\mathbf{B}[\mathbf{i},\mathbf{j}+\mathbf{1}]-\mathbf{B}[\mathbf{i},\mathbf{j}])^{\mathbf{2}}, \text{where}\ ^{-} > \mathbf{0} \tag{5.4}$$

This term *smoothes* **B**, and thus poses a penalty on variation in energy of appliance 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 **B** [102].

2. Appliance correlations: The energy usage of different appliances is often correlated [68]. 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 5.4.

5.6 Summary

Energy breakdown literature has largely looked at methods that require additional hardware to be installed. Due to prohibitive costs, it is unlikely that a significant proportion of the world will have access to such hardware. We presented a simple matrix factorisation based approach that does not require any sensing in the test home. 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.