Creating a Detailed Energy Breakdown from just the Monthly Electricity Bill

Nipun Batra
IIIT Delhi
Email: nipunb@iiitd.ac.in

Amarjeet Singh
IIIT Delhi
Email: homer@thesimpsons.com

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

***Abstract*—The first step to saving energy in the home is often to create an energy breakdown: the amount of energy used by each individual appliance in the home. Unfortunately, current techniques that produce an energy breakdown are not scalable: they require hardware to be installed in each and every home. In this paper, we propose a more scalable solution called Gemello that estimates the energy breakdown for one home by matching it with similar homes for which the breakdown is already known. This matching requires only the monthly energy bill and household characteristics such as square footage of the home and the size of the household. We evaluate this approach using 57 homes and results indicate that the accuracy of Gemello is comparable to or better than existing techniques that use sensing infrastructure in each home. The information required by Gemello is often publicly available and, as such, it can be immediately applied to many homes around the world.**

I. Introduction

Buildings account for more than 30% of total energy usage around the world, of which up to 93% is due to residential buildings [6]. Some of this energy can be saved by producing an *energy breakdown*: the amount of energy used by each individual appliance in the home, akin to the itemised bills we get from grocery stores. With such a breakdown, utility companies can focus conservation programs on homes that have an especially inefficient appliance, such as a fridge, or air conditioner. Energy feedback can also help induce energy-saving behaviour in the occupants themselves [5].

Unfortunately, current techniques that produce an energy breakdown are not *scalable*: they require hardware to be installed in each home and/or intensive manual training of the system. Non-intrusive load monitoring (NILM) tries to infer the energy breakdown based on the aggregate power trace from a single smart meter installed in the home. This approach is perhaps the most practical because smart meters are already being rolled out in millions of homes worldwide. However, current techniques often require high resolution data (1 minute sampling frequency or higher) [3], [9] while most smart meters today only support 15-minute or hourly sample rates to support time-of-use energy pricing. Even if smart meters had a higher sampling rate, most of the world does not yet have smart meters and many places do not even have plans to deploy them. Alternatives to NILM are more accurate but require specialized sensors to be installed inside the home. These solutions are limited by the need for instrumentation to be deployed in every home.

In this paper, we propose a more scalable solution called *Gemello* that produces an energy breakdown in homes without requiring new hardware to be installed in each home. Instead, Gemello estimates the energy breakdown for a home by matching it with similar homes that do have a hardware-based disaggregation solution. This matching requires only the monthly energy bill and household characteristics such as square footage of the home and the size of the household. From an energy perspective, homes in the same geographic region are often very similar because they have similar construction methods, use the same heating fuels, and contain similar fridges, washing machines, and other appliances. Gemello exploits this fact to provide an energy breakdown for many homes in a region by instrumenting only a fraction of them.

Of course, no two homes are exactly identical and finding a perfect twin is unlikely. Gemello uses a different set of matching homes to estimate the energy usage of each individual appliance. The key is the ability to define 'similarity' on a per-appliance basis. For example, homes with similar seasonal trends in their monthly energy bill are expected to have similar heating/cooling energy (referred as heating, ventilation, and air conditioning [HVAC] from now on); homes with similar square footage are expected to have similar lighting loads. The energy usage of each appliance is predicted by a different set of features, and so Gemello finds a different set of homes to predict the energy usage of each appliance.

We evaluate this approach using 57 homes from the publicly available Dataport data set [8], in which the ground truth energy breakdown is measured by metering each appliance of the home individually for one year or more. Results show that the accuracy of Gemello is comparable to or better than established NILM techniques called FHMM and LBM, both of which require on-site, high-frequency power metering in each home [7], [10]. Gemello can achieve 76% and 69% accuracy on HVAC and fridge, compared to 61% and 39% for FHMM and 56% and 72% for LBM. Furthermore, it achieves up to 57% accuracy on washing machine, dryer, dishwasher, and lighting loads, which is higher than previously reported results. Many existing techniques are not able to disaggregate these loads at all. The accuracy of Gemello becomes higher in homes with smart meters that provide power readings at 15-minute resolution. Gemello has potential for immediate impact because all of the information it requires is already available.

II. Approach

The goal of Gemello is to predict the energy consumption of household appliances given easy to collect information such as a single aggregate energy reading per month and static household characteristics. The key intuition behind Gemello is that if per appliance energy consumption is available for a

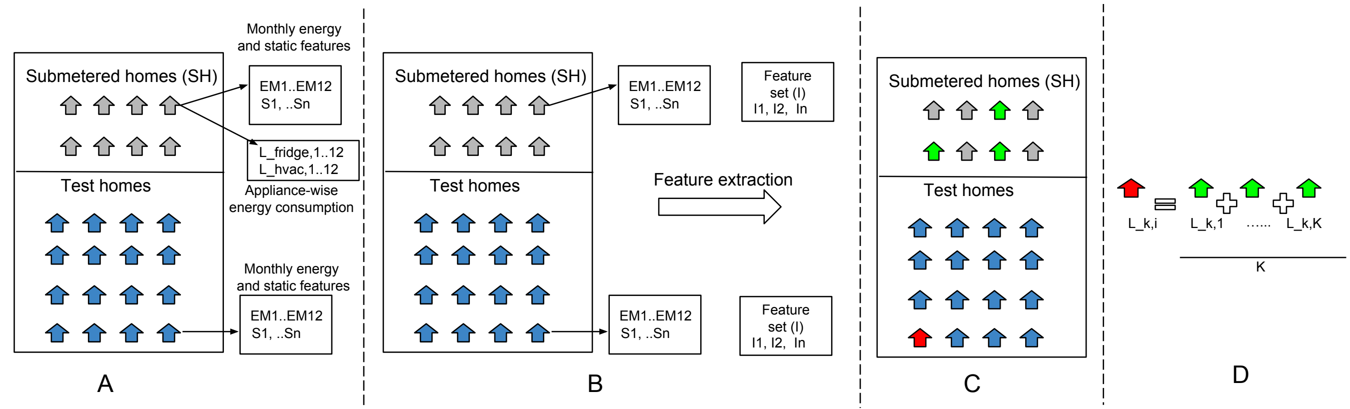

Fig. 1. Overview of Gemello. A) We assume a small set of submetered homes for which we have access to appliance level energy consumption. Both submetered homes and test homes have access to monthly energy and static household features; B) We extract features from the monthly and static household characteristics for all homes to obtain the feature set (I) per home; C) For a given test home (shown in red), for which we want to predict the energy consumption of a particular appliance, we choose the most similar homes (shown in green) for the test home. Similarity is defined on a per-appliance basis on the feature set (I). Each appliance will have different set of features for defining similarity (like number of occupants for washing machine, household area for lighting); D) We predict the appliance energy consumption of the test home as the average of the energy consumption of that appliance across the neighbours chosen in the previous step.

Feature category	**Feature sub category**	**Features (f)**	**#f**
Monthly energy	Raw monthly energy	12 month household energy aggregate	12
	Derived monthly energy	Variance over 12 month aggregate energy, Min. energy/Max. energy across 12 months, , Max. energy-Min. energy across 12 months, (Max. energy - Min. energy)/Max. energy, Skew, Kurtosis , 25^{th}, and 75^{th} percentile of 12 months energy	9
Static household characteristics		Household area, # occupants, # rooms	3
15 min AMI data	Time series decomposition	Max. of seasonal and trend component, Energy of seasonal component across months of HVAC usage , Top-7 Fourier transform coefficients of seasonal component	14
	Usage patterns	Autocorrelation (lag 1 day), Frac. of energy used across 24 hours, Frac. of energy used across 7 days	32
	Step changes	Proportion of step changes in 3 bins (<500 Watts, >500 and <1000 Watts; >1000 Watts), Cluster centers of above 3 bins	6

TABLE I
THESE FEATURES ARE USED TO DEFINE SIMILARITY BETWEEN HOMES. GEMELLO PRIMARILY USES FEATURES OF THE MONTHLY ENERGY BILL AND STATIC HOUSEHOLD CHARACTERISTICS. HOWEVER, IT CAN ALSO USE FEATURES OF 15-MINUTE AMI POWER DATA IF THE HOME HAS A SMART METER

home then any other home that is 'similar' in nature will also have the same energy consumption for that appliance. This intuition leads to the following requirements for Gemello:

1. Small set of submetered homes (SH) for which per appliance energy consumption is directly measured.

2. Easy to collect information for all the homes in the universal set U - This includes monthly energy consumption (j^{th} home represented by a vector $[EM_{j_1}, \ldots, EM_{j_{12}}]$) which is anyway collected for billing purposes and static household information such as size and number of occupants (represented by a vector $[S_{j_1}, S_{j_2}, \ldots]$).

For each target home i ($i \subset U$), we first derive the feature set (as discussed later) using vectors $[EM_{i_1}, \ldots, EM_{i_{12}}]$ and $[S_{i_1}, S_{i_2}, \ldots]$. Let these feature set be represented by $[I_{i_1}, I_{i_2}, \ldots]$. We next derive the K nearest neighbours for the i^{th} home from amongst the submetered homes in SH by comparing the feature set $[I_{i_1}, I_{i_2}, \ldots]$ with the feature sets derived for the remaining homes SH. Let this set of 'similar' homes be $[N_1, \ldots, N_K]$. Monthly energy consumption for kth appliance (L_k) for ith home is then calculated as:

$$L_{k_i} = \frac{L_{k_1} + L_{k_2} + \cdots + L_{k_K}}{K}$$

where $L_{k_1}, \ldots, L_{k_K}$ represent the monthly energy consumption for the same kth appliance from K identified 'neighbours' which are already sub-metered (and hence this information is directly measured). Figure 1 outlines Gemello.

We define 'similarity' differently for each electrical load. For example, homes having 'similar' monthly energy consumption are likely to have similar HVAC energy as the monthly energy consumption in many countries is often dominated by HVAC loads. Homes having 'similar' area are likely to have 'similar' lighting fixtures and thus may consume similar lighting energy. Homes having a 'similar' number of occupants may have 'similar' dish washer energy usage, as dishes loads is likely to be proportional to the number of occupants. By defining 'similarity' differently for each load, we leverage the fact that no two homes are exactly identical in all respects, but that every home is likely to have a set of similar homes that can be leveraged for predicting energy consumption at the load level.

For an electrical utility that has no smart metering infrastructure or homes in their customer base that already have appliance level monitors already installed, it can chose these set of homes in a way that all the remaining (large number of) homes are 'similar' to one or more of these neighbourhood homes. Standard clustering techniques can be used on the feature set derived from already available information, as is used in Gemello for estimating K neighbours, to create cluster of homes from their customer base and then a small set of

HVAC	Lights	Fridge	Dryer	Dish washer	Washing machine
31	12	21	32	26	16

TABLE II
HOMES IN THE DATASET WITH GROUND TRUTH VALUES

homes can be selected from each cluster for installing sub-metering infrastructure. We now discuss how 15-min. AMI data can be leveraged in Gemello.

Since 15-minute AMI data is very rich compared to monthly energy consumption, we divide the features derived from AMI data into the following broad classes:

I: Seasonal and Trend - To derive these features, we first decompose the 15-minute AMI data time series using Seasonal Decomposition of Time Series by Loess (STL) [4] into seasonal and a trend 15-minute timeseries. Features derived in this class include max., energy and first 7 coefficients obtained from the Fourier transform of the seasonal component.

II: Temporal Patterns - Features derived in this class include fraction of energy usage across the 7 days of the week, energy usage across the 24 hours of the day and energy usage during the night time. Such features can be used to derive energy consumption of appliances showing temporal usage patterns. For example, certain working class populations may do their groceries and dishes on weekends and thus would have higher fridge and dryer energy consumption on weekends.

III: Step changes - A step change is likely to be caused by an appliance changing its state. For example, washing machine turning ON could cause a step change of 500 Watts. We compute the fraction of step changes whose magnitude are less than 500 Watts, between 500 and 1000 Watts and greater than 1000 Watts. Step changes greater than 1000 Watts are caused by high power appliances such as HVAC, while step changes less than 500 Watts are caused by low power appliances.

The total set of features is summarised in Table I. Since these features are on different scales, we normalise them in the range 0 to 1 using standard scaling procedures.

III. Evaluation

A. Data set and Baseline approaches

We use the publicly available Dataport data set [8] for evaluation. 57 homes contain data for 1 year (2013) across 6 appliances of interest (fridge, HVAC, lights, dryer, washing machine and dish washer) and house metadata such as the number of occupants, area of the home and number of rooms. While the data set contains submetered data from a large number of appliances, only the chosen 6 appliances had data across significant number of homes. Not all of these 57 homes monitored all these 6 appliances. Table II shows the number of homes for each of these 6 appliance types.

We compare the accuracy of our approach against the following two approaches: Factorial Hidden Markov Model (FHMM) [7] and Latent Bayesian Melding [10].

B. Evaluation metric

We define our metric based on prior work [1], [7]. Let the predicted and ground truth appliance energy for an appliance a in home h for month m be $Pred(h, a, m)$ and $GT(h, a, m)$ respectively. The absolute percentage error (PE) for the h, a, m triplet is given by:

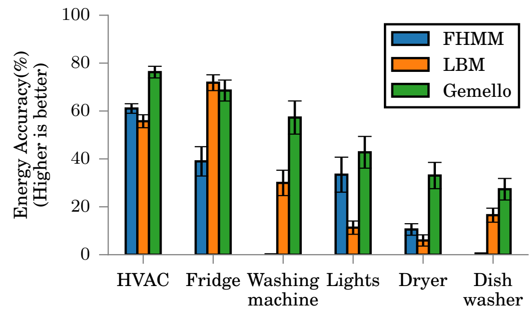

Fig. 2. Gemello achieves Energy Accuracy that is comparable to or better than the baseline approaches across all appliances

$$\mathrm{PE(h,a,m)} = \frac{|\mathrm{Pred(h,a,m)} - \mathrm{GT(h,a,m)}|}{\mathrm{GT(h,a,m)}} \times 100\%$$

The percentage accuracy (PA) for h, a, m triplet is given by:

$$\mathrm{PA(h,a,m)} = \begin{cases} 100 - \mathrm{PE(h,a,m)}, & \text{if } \mathrm{PE(h,a,m)} \leq 100 \\ 0, & \text{otherwise} \end{cases}$$

The percentage accuracy ($PA(h, a)$) for an appliance a in home h is given by the mean of $PA(h, a, m)$ across all months.

$$PA(h, a) = \overline{PA(h, a, m)}$$

We now define our main metric Energy accuracy (%) for an appliance a to be the mean of $PA(h, a)$ across all homes.

$$\text{Energy accuracy}(a) = \overline{PA(h, a)}$$

In our results, we also calculate the standard error to show the variation in energy accuracy across homes. Higher 'Energy accuracy' indicates better disaggregation performance.

C. Experimental setup

Our experimental setup tries to replicate the following real world scenario- we have a small subset of submetered homes and a large number of homes without smart meters. However, our data set has the limitation that only a small number of homes (Table II) containing aggregate and appliance energy for long duration (1 year) and household characteristics is available. Thus, we use the *leave-one-out* cross validation technique to evaluate our approach, where we assume the set of submetered homes to be all the homes except the test home. Since we also want to tune the parameters of our model, we do a nested cross-validation, where the inner loop of cross validation is used for model tuning. We tune our model on two parameters- the number of neighbours (K) and the top-N features. We vary K from 1 to 6 and N from 1 to max(number of features, 10). The top-N features are learnt using the ExtraTreeRegressor using the default parameters provided in the Scikit-learn implementation. We use data from 2013 for our evaluation. For all homes, we had aggregate and appliance monthly energy consumption for the 12 months. For the HVAC, the evaluation was done only on the months (May-October) in which it is typically used in Texas.

To compare our performance with FHMM, we train an FHMM on 15-minute appliance level data from all homes across the dataset using nilmtk [1]. Each appliance is modelled as a 3 state HMM, as is commonly done in the literature [1], [10]. The FHMM is composed of an HMM for each home

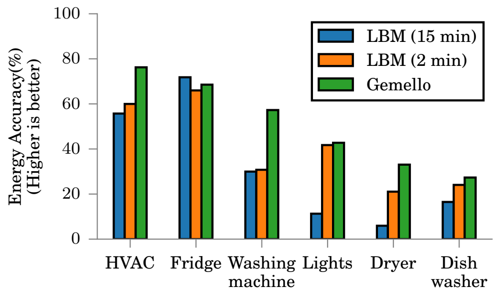

Fig. 3. Gemello achieves higher accuracy with monthly sampling frequency than the state-of-the-art LBM algorithm achieves with 15-minute or even 2-minute sampling frequency.

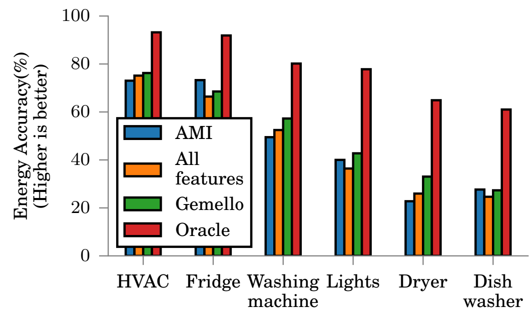

Fig. 4. The performance of Gemello can be improved by up to 5% for some appliances if 15-minute AMI data is available. The Oracle illustrates the upper bound on accuracy that can be achieved with this data set if the neighbour selection algorithm were perfect.

contains the 6 appliances of interest. It must be noted that the result of FHMM disaggregation is a 15-minute power signal for each appliance. We calculate the per month energy from this power signal. To compare our performance with LBM, we use the implementation provided by the paper authors. The FHMM is trained on 15-minute appliance level data from the dataset. Population statistics such as appliance number of cycles, duration of usage are computed on the entire dataset. The LBM paper authors confirmed that the parameters and hyper-parameters we learnt were reasonable. Like FHMM, LBM outputs a 15-minute power timeseries, from which we calculate the monthly energy consumption per appliance.

D. Results and analysis

Our main result in Figure 2 shows that Gemello gives comparable or better disaggregation performance across all loads in comparison to the 2 baseline approaches. Gemello can achieve 76% and 69% accuracy on HVAC and fridge, compared to 61% and 39% for FHMM and 56% and 72% for LBM. Furthermore, it achieves up to 57% accuracy on washing machine, dryer, dishwasher, and lighting loads, which is higher than previously reported results. Many existing techniques are not able to disaggregate these loads at all. We discussed the results of the benchmark algorithms with several energy disaggregation researchers and found that there are several reasons why the performance of FHMM and LBM is poor in comparison to our approach, such as FHMM is known to work poorly when the 'unexplained' energy is high.

We evaluated the performance of our approach in comparison to LBM in the case higher resolution data was available to LBM. In Figure 3, we see that with an increase in resolution from 15 to 2 minutes, the accuracy of LBM improves. However, our approach still performs comparable or better across all appliances.

Having established that Gemello gives better or comparable accuracy than state-of-the-art algorithms, we now analyse the performance of Gemello if additional data such as 15-minute AMI is also available. Figure 4 shows the performance of Gemello across different feature sets. 'All' set of features contains features from AMI data in addition to static household and monthly features. For fridge, the AMI features increase accuracy by 5%. On analysing the optimal set of features selected during nested cross validation when fed AMI features, we found that in 20/21 homes, proportion of step changes that are less than 500 Watts was among the optimal features. As we had hypothesised earlier, fridge is a duty cycled appliance often consuming less than 500 Watts. Thus, homes similar in proportion of step changes less than 500 Watts are similar in fridge consumption. We also found the fraction of energy consumed in afternoon hours to be an important feature as it occurs in 19/21 homes. Homes show low electrical activity during afternoon hours, which is often dominated by the always-on appliance fridge. Thus, fraction of energy in afternoon hours is useful for finding homes with similar fridge energy. To find the best possible performance of Gemello on our dataset, we define the 'Oracle'. The 'Oracle' finds the 'optimal' subset of 'neighbourhood' homes from all possible sub sets of homes for each home for each load. The appliance energy data from these 'optimal' homes is then averaged to predict the appliance energy consumption for the test home. Here, 'optimality' is defined as giving the best energy breakdown accuracy. The substantially high accuracy of the 'Oracle' only highlights the potential of Gemello . If we were to find *better* features for defining similarity between homes, our accuracies would tend towards the accuracy of the 'Oracle'.

IV. Conclusions

A detailed energy breakdown is the first step towards saving energy in a home. We started with the goal of creating an energy breakdown solution that - works with whatever data is easily accessible, is able to scale across large number of homes and requires minimal capital expenditure involved. In order to achieve these objectives, we completely flipped the way we look at the problem. Rather than the existing bottom-up approach of using modelling to identify electrical signatures, we used the top-down approach of using modelling to identify home level characteristics that correlate well with appliance level energy consumption. We show that such home level characteristics can be easily calculated with static household information and monthly electricity data both of which are readily available. Results show that our proposed approach that satisfies the objectives we started off with, gives comparable or better energy breakdown accuracy than the state-of-the-art NILM techniques which rely on high resolution power consumption data from smart meters. We believe that Gemello has potential for immediate impact as all the information it requires is already available.

V. Note

This work is an abridged version of the following report [2].

REFERENCES

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

[2] N. Batra, A. Singh, and K. Whitehouse. Gemello: Creating a detailed energy breakdown from just the monthly electricity bill. 2016.

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

[4] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–73, 1990.

[5] S. Darby. The effectiveness of feedback on energy consumption. *A Review for DEFRA of the Literature on Metering, Billing and direct Displays*, 2006.

[6] M. Evans, B. Shui, and S. Somasundaram. *Country report on building energy codes in india*. Pacific Northwest National Laboratory, 2009.

[7] J. Z. Kolter and T. Jaakkola. Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pages 1472–1482, La Palma, Canary Islands, 2012.

[8] 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. 2015.

[9] O. Parson, S. Ghosh, M. Weal, and A. Rogers. Non-intrusive load monitoring using prior models of general appliance types. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, pages 356–362, Toronto, ON, Canada, 2012.

[10] M. Zhong, N. Goddard, and C. Sutton. Latent bayesian melding for integrating individual and population models. In *Advances in Neural Information Processing Systems 28*, pages 3617–3625. 2015.