Collaborative filtering for household load prediction given contextual information

Taha Hasan*, Naveed Arshad†, Erik Dahlquist‡, Scott McCrickard*

Abstract

Matrix factorization models express the relationships between energy demand profiles of a collection of households as inner products in a joint latent factor space inferred from known consumption data. Neighborhood models explicitly characterize the said relationships by determining how similar the consumption patterns are among these various households. This study leverages the two strategies to describe a collaborative filtering framework for the individual household load prediction problem. It provides the potential to incorporate weather, structural and sociological data and improve on the explainability and accuracy of constituent models.

1 Introduction

Household short-term load forecasting is a key enabling technology for several envisioned functionalities of the comprehensive residential energy management system (EMS). A forecast of state variables (weather, temperature, local demand/generation) is central to automated scheduling and control of household appliances in systems like Energy Box [11]. Device-level and aggregate load 'profiles' for households help identify and visualize patterns of energy use for consumers and researchers [12]. PHEV integration in household microgrids, vehicle-to-grid technology and demand response (DR) services require careful modeling of expected household demand [7]. However, there is a marked increase in time-series volatility for household load data (up to several orders of magnitude) relative to aggregate load for a colony, microgrid or a power-system at large [6], [5]. This suggests the need for forecasting solutions that are robust to electrical noise, load and seasonal dynamics. It also emphasizes the need to leverage structural and contextual information to capture usage patterns local to a sub-group of households or to a significant period of usage and the relationships between the two.

Collaborative filtering models are among the most competitive and intuitive models for identifying and leveraging local relationships in sparse matrices and generating concise representations thereof [9]. This study describes a collaborative filtering approach to predict energy demand profiles for a collection of households given contextual information.

*Computer Science Department, Virginia Tech

†Computer Science Department, Lahore University of Management Sciences

‡Public Technology Department, Malardalen University

2 Problem statement

Let X_h be the data matrix corresponding to energy usage of a single household worth N_w weeks sampled at R_s hr^{-1}. $X \in \mathbb{R}^{N_w \times N_s}$, $N_s = 168 \times R_s$. Weather and humidity data matrices W and H are of same dimensionality. Matrix entries for test weeks need to be estimated for all N_h households.

2.1 Global effects and biases Global effects and biases model the enduring effects of each household and each specific time interval in $P_{h,t}$ independent of their interaction. Baseline consumption for each household in the SVD model ($P_{h,t} = \mu + b_h + b_t$)[9] identifies the global bias μ, the variance contributed by each household h, b_h and by each timestamp t, b_t. A couple of aspects are important to discuss here. One, high volatility/low SNR renders the global bias ineffective and leads to poorer precision of prediction on the household scale relative to that on the colony/neighborhood scale. Two, accounting for the variance in $P_{h,t}$ from each timestamp leads to a substantially large number of features and models the noise and transient effects in addition to the signal of interest. Instead, a fixed number of features - corresponding to hour, day and month indices and average outdoor temperature etc. - are used as discriminants.

$$P_{h,t} = b_h + \sum_k b_{d,k} \tag{2.1}$$

A least squares problem can be devised to learn the parameters b_h and $b_{d1..k}$ as follows.

$$\min_b \sum_h \sum_t (y_{h,t} - b_h - b_{d1..k})^2 + \lambda_1 * (\sum_h b_h^2 + \sum_k b_{d,k}^2)$$

The first term minimizes the squared error between measured power usage at a given time instant for each household and the baseline model. The second term penalizes the magnitude of the various terms in the baseline model to prevent overfitting and keep the predictions generalizable.

2.2 Matrix factorization models Matrix factorization models infer a latent-factor space which expresses a co-clustering interpretation [8] for electricity consumption data using inner products in the factor space [9]. The factors q_t and p_h are d-dimensional tuples and represent the timestamp t and household h in the d-dimensional factor space, respectively. Each household-timestamp pair is, in turn, represented as the inner product $q_t^T p_h$.

$$P_{h,t} = q_t^T p_h,\ q, p \in \mathbb{R}^d \tag{2.2}$$

The model parameters are learnt by solving the following optimization problem.

$$\min_{p,q} \sum_h \sum_t (y_{h,t} - q_t^T p_h)^2 + \lambda_2 * (\sum_h ||p_h||^2 + \sum_t ||q_t||^2)$$

An alternating least squares (ALS) approach [2], [3] or stochastic gradient descent [9] can be used to learn these parameters.

2.3 Neighborhood models Surveyed literature for this study brings the regression performance of neighborhood modeling into question. In [5], proximity in time-series cross-correlation identifies k households that 'lead' the energy consumption of a given household. Prior consumption data (previous hour/day) from these households is incorporated in the data matrix for each household, albeit with no significant gains in prediction accuracy for regression or kernel-based models. Such a model marginally compensates accuracy for the case where previous hour data fails to capture sharp spikes and swells in residential demand. On the other hand, predictive power is also favored by including consumption data - for all households - and contextual data - for each individual household - directly in the data matrix, as demonstrated in [6]. Avoiding the segmentation of customer data in this fashion, however, arguably loses the explainability of resulting models. Experiments in section 4 suggest that an improved accuracy of prediction using recent historical data of k nearest neighbors requires careful feature selection and regularization.

The neighborhood model in eq. 2.3 samples the previous day's consumption data of k-th neighbor weighted by a similarity index s_{hk}. A regularization constant α ensures that neighborhood information for distant households is not captured by the model.

Eqs. 2.2 and 2.3 are combined to deliver the final prediction rule in eq. 2.4.

$$P_{h,t} = \gamma_3 \frac{\sum_k s_{hk} P_{k,prevday}}{\alpha + \sum_k s_{hk}} \tag{2.3}$$

$$P_{h,t} = b_h + b_{d1..k} + q_t^T p_h + \gamma_4 \frac{\sum_k s_{hk} P_{k,prevday}}{\alpha + \sum_k s_{hk}} \tag{2.4}$$

All parameters for eq. 2.4 are simultaneously learnt in the pattern of eqs. 2.1 and 2.2 by minimizing the following objective function.

$$\min_{b,p,q} \sum_h \sum_t (y_{h,t} - b_h - b_{d1..k} - q_t^T p_h - \gamma_4 \frac{\sum_k s_{hk} P_{k,prevday}}{\alpha + \sum_k s_{hk}})^2$$
$$+ \lambda_4 * (\sum_h (b_h^2 + ||p_h||^2) + \sum_k b_{dk}^2 + \sum_t ||q_t||^2 + \gamma_4^2)$$

2.4 Prediction strategies For day-ahead load prediction, unknown rows of the data matrix are estimated using day selection heuristics described in PJM's analysis of DR baselines[10]. In general, long-term prediction can be carried out by asserting a 'context' for each test day [4] using criteria such as same hour/month/day indices, proximity in weather conditions or k-previous days. The factor products for same-context training days are averaged for each test day. Section 4 evaluates these criteria for day-ahead prediction.

As an alternative to the use of empirical day selection heuristics, a kernelized extension of the sparse factor model can be pursued. Authors in [8] cast the load data matrix X as sum of a low-rank component L and a sparse factor model $Q^T P$. Elements of L and $Q^T P$ are parametrized in terms of kernel matrices K_{l1}, K_{l2}, K_q and K_p which encode information about time-series correlation with prior weeks. Block coordinate descent (BCD) is employed to iteratively compute the constituent blocks of L and $Q^T P$. For future work, we intend to incorporate weather and contextual data directly in the product kernel.

3 Evaluation

3.1 Dataset and Initial Conditions Dataset for this study is contributed by a 2007 project initiated by Ireland's Commission for Energy Regulation (CER) to assess the behavioral impact of energy usage feedback on household electricity consumers. The project conducted in-depth customer behaviour trials (CBTs) during 2009 and 2010 with over 5,000 Irish homes and businesses across a variety of demographics and home sizes [1] contributing consumption, structural and contextual data.

Table 1: ISSDA pre-trial survey: sample questions

ID	Sociological	ID	Structural
200/300	Respondent's age/sex	450	Apt./detached/semi-detached
310	Employment status of chief income earner	453/460/6103	House build year/# bedrooms/floor area
420/431	Residents above/below 15 years of age	470	Heating: central/plug-in/gas/oil/renewable
4021	Yearly household income before tax	4704	Cooking: electric/gas/oil/solid fuel
4331X	Motivation: reducing carbon footprint	4900X	# appliances: washer, dryer, dishwasher
4331X	Motivation: saving money on utilities	4900X	# appliances: electric heater, cooker, PC
4352	Prior knowledge of power usage of appliances	4900X	# hours daily usage for each appliance

Table 2: Effect of neighbors' consumption data ($k = 32$) on regression error

Alg.	MLR (Corr)	MLR (Eu)	LASSO (Eu)	LASSO (Cos)
ΔMAPE (test, %)	+2.9	+2.0	-4.3	-4.1

The dataset includes pre-trial and post-trial questionnaire data examining the structural aspects of the household, respondents' financial background, energy awareness and initiative for savings. Sample questions from the pre-trial survey are listed in table 1. The consumption data includes aggregate electricity and natural gas usage sampled every half-hour for participating households over a period of 18 months. Only control group houses (i.e. no identified change in consumption patterns with feedback) are included in the analysis. Described models are trained on first twelve months of the dataset and predictor performance is evaluated on the final six. A moving average filter (with a six-hour window) is applied to the dataset to improve the signal-to-noise ratio and remove null patches.

3.2 Use of household similarity LASSO regression and a dissimilarity matrix computed using Euclidean or cosine distance goes some way towards reducing MAPE with an increase in k, the number of neighbor households.

3.3 Day-selection heuristics PJM's analysis of customer baseline protocols [10] provides a variety of data-selection rules addressing criteria such as proximity to event day in time, power usage and weather conditions. A subset of these rules (PJM Economic/GLD/GLD Weather-sensitive, NYISO, CAISO) is evaluated for inferring day-ahead electricity demand in our model. Table 3a specifies the selected days for each protocol. The first three protocols (PJM Economic, CAISO, NYISO) in table 3a are of 'highest X of Y' type (X highest kWh days out of Y most recent calendar days chosen from Y' candidate days). PJM GLD (weather-sensitive) ranks candidate days by difference in temperature-humidity index (THI) relative to event day and chooses the minimum rank day for comparison. Day-ahead household demand is predicted by averaging the inner products for days ranked closest to that day by described rules. Table 3b describes the accuracy of prediction for these protocols. CAISO offers the minimum RMSE and uses the largest number of days in the evaluated rules. Prediction accuracy for CAISO declines as the number of days accounted for in the comparison exceeds 30.

3.4 Results Table 4 lists the prediction error for day-ahead forecasts using described baseline models (Prev-week, All-days and PJM-WS) and collaborative filtering models in eqs.2.1-2.3 (referred in table 4 as M_1-M_3).

4 Discussion

It is important to comment on the administrative cost of baseline estimation methods. An increase in the number of candidate days factored towards an estimation method implies a similar increase in cost of data transfer, storage and analysis for market participants [10]. The kernelized extension of M_2 might help avoid the use of empirical day selection strategies,however, the use of usage data from k-previous weeks for estimation of kernel matrices warrants an operational feasibility analysis.

References

[1] Irish Social Science Data Archive - CER, 2015.

[2] Robert Bell, Yehuda Koren, and Chris Volinsky. Modeling relationships at multiple scales to improve accuracy of large recommender systems. In *Proceedings of*

Table 3: CBL protocols: data selection rules and accuracy of prediction

(a) Data selection rules [10]

CBL Protocol	Day Selection Rules
PJM Economic	$Y' = 45(+15)$
	Weekdays: $X = 4$, $Y = 5$
	Weekends/holidays: $X = 2$, $Y = 3$
CAISO Standard	Weekdays: $X, Y = 10$
	Weekends/holidays: $X, Y = 4$
NYISO Standard	$Y' = 10$
	Weekdays: $X = 5$, $Y = 10$
	Weekends: $X = 2$, $Y = 3$
PJM Emergency GLD	Weekday only, closest non-holiday weekday
PJM GLD (Weather-sensitive)	Weekday only, day with lowest difference in temperature-humidity index (THI) relative to event day

(b) Prediction error using the latent factor approach (M_2)

Data selection strategy	all-days	prev-day	NYISO	PJM-ECO	CAISO	PJM-GLD	PJM-WS
RMSE	0.7253	0.6663	0.6671	0.6646	0.6614	0.7682	0.8378

Table 4: Prediction error (test) for baselines and proposed CF models in Section 3

Alg.	PJM-WS	Prev-day/all-day	M_1(baseline)	M_3	M_2
RMSE	0.96	0.77/0.70	0.95	0.74	0.66
NRMSE	1.06	0.86/0.89	1.35	0.89	0.76

the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 95–104. ACM, 2007.

[3] Michael W Berry, Murray Browne, Amy N Langville, V Paul Pauca, and Robert J Plemmons. Algorithms and applications for approximate nonnegative matrix factorization. *Computational statistics & data analysis*, 52(1):155–173, 2007.

[4] Vikas Chandan, Tanuja Ganu, Tri Kurniawan Wijaya, Marilena Minou, George Stamoulis, George Thanos, and Deva P Seetharam. iDR: consumer and grid friendly demand response system. In *Proceedings of the 5th international conference on Future energy systems*, pages 183–194. ACM, 2014.

[5] S. Humeau, TK. Wijaya, M. Vasirani, and K. Aberer. Electricity load forecasting for residential customers: exploiting aggregation and correlation between households. In *Sustainable Internet and ICT for Sustainability (SustainIT)*, 2013.

[6] F. Javed, N. Arshad, F. Wallin, I. Vassileva, and E. Dahlquist. Forecasting for demand response in smart grids: An analysis on use of anthropologic and structural data and short term multiple loads forecasting. *Applied Energy*, 96:150–160, 2012.

[7] Linni Jian, Honghong Xue, Guoqing Xu, Xinyu Zhu, Dongfang Zhao, and Z.Y. Shao. Regulated charging of plug-in hybrid electric vehicles for minimizing load variance in household smart microgrid. *Industrial Electronics, IEEE Transactions on*, 60(8):3218–3226, Aug 2013.

[8] SJ. Kim and GB. Giannakis. Load forecasting via low rank plus sparse matrix factorization. In *Proceedings of the 2013 Asilomar Conference on Signals, Systems and Computers,*, 2013.

[9] Yehuda Koren and Robert Bell. Advances in collaborative filtering. In Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, editors, *Recommender Systems Handbook*, pages 145–186. Springer US, 2011.

[10] P. Langbein and T. Hennessy. PJM Empirical Analysis of Demand Response Baseline Methods. Technical report, PJM Inc./KEMA, 2011.

[11] Daniel Livengood and Richard Larson. The Energy Box: Locally Automated Optimal Control of Residential Electricity Usage. *Service Science*, 1(1):1–16, 2009.

[12] Joakim Widn, Magdalena Lundh, Iana Vassileva, Erik Dahlquist, Kajsa Ellegrd, and Ewa Wckelgrd. Constructing load profiles for household electricity and hot water from time-use datamodelling approach and validation. *Energy and Buildings*, 41(7):753 – 768, 2009.