Energy Aggregation using Product of HMMs

Megha Gupta
Department of Computer Science
IIIT Delhi, India
meghag@iiitd.ac.in

Haimonti Dutta*
Department of Management Science and Systems,
State University of New York, Buffalo
New York, 14260
haimonti@buffalo.edu

Ullas Nambiar
EMC Corporation
Bangalore, India
ullas.nambiar@emc.com

Amarjeet Singh
Department of Computer Science
IIIT Delhi
amarjeet@iiitd.ac.in

ABSTRACT

The need to gather fine grained real-time spatio-temporal energy consumption data is fulfilled by the large scale deployment of smart meters. Remote monitoring on these meters is done by sending readings from the customer site to the data aggregators placed at the substations. Each substation aggregates the load derived from all the meters connected to that substation. The readings received at the substation are adhoc and usually not synchronized in time. Different smart meters can send data points when they are collected resulting in inconsistent data including aggregating non-aligned time stamped readings, readings with missing values, repeated values, meter reset readings. We address the problem of learning from disparate data streams (with inconsistencies) by modelling streams as HMMs and the process of aggregating data at the substation as a Product of HMMs. This enables us to perform load forecasting using machine learning techniques. We have performed experiments using contrastive divergence learning on the REDD data set and the energy consumption data collected from the faculty housing at our institute. The results show that this technique performs best by combining larger no. of HMMs with smaller no. of states (on, off or standby) with training time linear to the number of HMMs. via a product.

Keywords

Energy aggregation; Ensemble learning; Product of HMMs

*The author is also affiliated to the Institute of Data Science and Engineering (IDSE), Columbia University and is an adjunct professor at IIIT-Delhi.

1. INTRODUCTION

Smart meters consisting of real time sensors, power outage notifications and power quality monitoring are widely used today. These meters provide a host of benefits like energy efficiency and savings, improved retail competition, better demand response actions, improved tariffs, lower bills due to better customer feedback, accurate billing, less environmental pollution, etc. They generate huge amount of data which helps in giving meaningful insights through analytics. They can measure site specific information and also help agencies to set different electricity prices for consumption based on the time of the day, seasons, holidays, etc. As a result, a feedback is sent to the customers by the utilities that can help consumers better manage their resources. A research [21] shows that by providing real time feedback, consumers can reduce the consumption by 3-5%. Also, for some country providing real time feedback may not be a cost effective plan but it can help in retaining customers and peak-load shift.

In recent years, machine learning has been applied to the problem of energy consumption and demand forecasting analysis. The role of the machine learning algorithm is to study the sensor data and provide alerts and warnings when anomalous behaviour occurs or to inform (and remind) customers when certain activities were performed, which rooms they occupied, and what appliances they used most frequently during that period. This information can be transmitted to customers in timely fashion via phone, email or the Internet. This paper [7] does a comparison of several clustering techniques and finds out that the hierarchical clustering and modified follow-the-leader perform best among the rest K-Means, fuzzy K-Means to group customers with similar electrical behaviour [20]. Another paper [28] uses classifiers like random forest, J48, logistic and naive bayes to identify customers with similar electricity consumption profiles. Sensor data collected from smart homes are used to reveal activity patterns of the residents, which can then be correlated with the total energy consumption. This enables utility companies and their customers to associate activities with energy usage and costs, devise intelligent systems to control home environments improving energy efficiency and reducing costs. Typically, sequences of usage patterns that appear frequently at different time scales (daily, weekly, monthly, yearly) and across different homes are studied and outlier detection algorithms are designed to enable customers to be notified that they are consuming unusually large amount(s) of energy during some specific period. Related problems involve study of trends of electricity consumption (steadily increasing, decreasing, cyclic, seasonal) and sudden anomalous behaviour (sudden peaks or drops on consumption) for individual homes and across the community.

In this paper, we use HMMs to model time series data. We build machine learning models using products of HMMs and apply them to the energy aggregation problem. Two different proof of concepts are presented – first one on the REDD data set and the other one on real data collected at the faculty housing in India. There are many reasons why the product model constructed from many HMMs is appropriate. First, this model is ideal for data which is caused by multiple underlying influences. Second, HMMs alone are not efficient at capturing long range structure in time series [27] – in contrast to product of hidden markov models (PoHMM) [5] allow each model to remember a different piece of information about the past.

Organization: This paper is organized as follows: Section 2 examines related work on data analytics on aggregated data of smart meters; Section 3 provides a review of products of Hidden Markov Models (HMMs) and how they relate to our application. The two proofs of concepts are introduced in Section 4 to illustrate the effectiveness of the use of product of HMMs in the energy aggregation problem. Finally, Section **??** concludes the work.

2. RELATED WORK

In this section, we describe work that uses ensemble learning techniques and non-ensemble learning techniques to solve problems in energy domain.

2.1 Non-ensemble based learning techniques

2.1.1 Energy Aggregation

In wireless sensor networks, energy data aggregation is a method of combining data from different sources and expressing on a specific variable, in a summarised format. As the sensor network generates lot of data for the end user to process, there are automated methods employed to aggregate data. This data fusion is generally known as data aggregation which combines the data into a set of meaningful information [15]. The sensor nodes are organised in a tree structure, called aggregation tree. The leaves of this tree are the sensor devices, the internal nodes are the aggregator devices that takes the data from the leaves, aggregates it and sends it to its parent node which is the root of the tree.

The main objective of data aggregation is to reduce the unnecessary information thereby reducing the network traffic and improving the privacy of the customers from internal and external entities by keeping only the necessary information [26].

2.1.2 Energy Disaggregation

The process in which the whole building energy (aggregated) signal is separated into appliance level energy (disaggregated) for a variety of reasons like residential energy reductions, program evaluation, targeted marketing, etc. Several studies have been done in this regard, one of the unsupervised desegregation method [17] that outperforms other unsupervised disaggregation methods is conditional factorial hidden semi-Markov model. This model when integrated with other features, accurately represents the individual appliance energy consumption. Another research [19] that exploits the additive structure of the FHMM to develop approximate inference procedure in energy disaggregation domain that outperforms the rest.

2.1.3 Load Forecasting

Electrical load forecasting refers to the projection of electrical load required in a certain geographical area with the use of previous electrical load usage in the same area. It is extremely important for efficient power system planning and operation, energy purchasing and generation, load switching, infrastructure development. It encompasses various factors like, historical load, weather data, population, energy supply and price, time of the year, etc. It is usually divided into three categories, short-term forecasts (one hour to one week) , medium-term forecasts (one week to one year) and long-term forecasts (more than a year). In short term load forecast, [2] and [6] used a three layer feed forward artificial neural network and to predict daily load profiles. In a paper by [8], nonlinear autoregressive integrated neural network was used to predict daily load consumption. In medium term load forecasts, the author forecasts [11] the monthly load through knowledge based activities from the output of the ANN based stage providing yearly energy predictions. Whereas in [3], time lagged feedforward neural network is used to do monthly forecasting on the basis of historical series of electrical load, economic and demographic variables. And the authors from covenant university, [23] performed load forecasting of their own educational institute using the models based on linear, compound growth and cubic methods of regression analysis. In long term load forecasting, study done by [9] resulted in showing that the models based on regression analysis did not give very accurate predictions as compared to fuzzy neural network which performed better due to better handling with non linear systems. Another work [30] uses support vector regression to derive non linear relationship between load and economic factors like GDP for long term forecasting in developing countries.

2.1.4 Customer Segmentation

The identification of consumer profiles that show similar behaviour in energy consumption. This analysis is useful in various ways, like demand response system, intelligent distribution channel. The author [29] segments the customers based on contextual dimensions like location, seasons, weather patterns, holidays, etc which help with various higher level applications like usage-specific tariff structure, theft detection, etc. In [1], author proposes to infer occupancy states from consumption time series data by using HMM framework. They investigate the effectiveness of HMM and model based cluster analysis in producing meaningful features of the classification. This work suggests the dynamics of time series as captured by HMM analysis can be valuable.

2.2 Ensemble based learning techniques

Ensemble learning is a method where multiple learners are trained to solve the same problem. It constructs a set of hypothesis and combines them to generate the final result.

2.2.1 Prediction with expert advice

A study done by [24], proposes a Pattern Forecasting Ensemble Model (PFEM) comprising of five forecasting models using different clustering techniques, like k-means model, self-organising map model, hierarchical clustering model, k-medoids model and fuzzy c-means model. They have showed that on three real-world dataset, their proposed ensemble model outperformed all the five individual model in case of day ahead electricity demand prediction. Another study [12] highlights the importance of regularised negative correlation learning ensemble methodology on the problem of energy load hourly prediction. This method tries to overcomes the problem of variability in neural network due to high sensitivitiness to the initial conditions. As this method combines the outputs of several neural networks, it achieves a marked reduction in error after introducing external data. In our paper, we deal with the problem of energy aggregation using ensemble learning model. Each HMM is used to represent a state of an appliance. An appliance can have states like ON or OFF. The combination of the outputs from each of these HMM models gives us our ensemble based learning model, Product of Hidden Markov Model (PoHMM) [16]. This learning technique outputs the probability distribution by combining the outputs from several simpler distributions. It allows each model to make a decision on the basis of few dimensions.

3. REVIEW OF POHMM

A hidden Markov Model (HMM) is a tool that represents the probability distribution over the sequence of observations [13]. It holds two name defining properties, first, the observation at time t is generated by a process whose state S_t is hidden from the observer and second, is that this hidden state process satisfies Markov property which states that given the value at state S_{t-1}, the value at current state S_t is independent of all the states prior to $t-1$. Another assumption that it follows is that the hidden state variable is discreet, that is S_t can take K values denoted by integers {1,..K}. In order to define probably distribution over the sequence of observation, it is important to define probability distribution over the initial state P(S_1), the transition probability P($S_t|S_{t-1}$) and the observed probability P($Y_t|S_t$) where Y_t is the observation at time t.

Using the traditional HMM notation for the parameter λ = {A, B, π } where A is the transition probability, B is the observed probability and π is the initial state probability. For HMMs, S^1 & S^2 we have the values of A, B, π as shown in table 1, 2, 3 respectively.

A_{S^1}	S^1_1	S^1_2
R_1	0.6	0.4
R_2	0.3	0.7

A_{S^2}	S^2_1	S^2_2	S^2_3
S_1	0.6	0.3	0.1
S_2	0.4	0.1	0.5
S_3	0.2	0.4	0.4

Table 1: Transition probabilities, A

3.1 Product of HMM

PoHMM is a way of combining multiple HMMs by multiplying their individual distribution together and then renormalizing them. Its representation includes both directed and undirected links where the hidden states are causally connected to the other hidden states but non causally related to the visible states. This causes different conditional independence relationships among the variables in graphical model. The figure **??** is a product of two HMMs shown in **??**. For P = S^1 x S^2, the number of states is the product of states in S^1 and S^2 which is 6. The connections formed in the P depend on the links in the multiplying HMMs. The resultant HMM will have a pair (s,s) X = {6; R_1S_1, R_1S_2, R_1S_3, R_2S_1, R_2S_2, R_2S_3}
X_0 = {R_1S_1}
A = {a,b}

B_{S^1}	a	b
S^1_1	0.2	0.8
S^1_2	0.5	0.5

B_{S^2}	a	b	c
S^2_1	0.2	0.3	0.5
S^2_2	0.5	0.4	0.1
S^2_3	0.4	0.3	0.3

Table 2: Observed probabilities, B

	R_1	R_2
π_{S^1}	0.4	0.6

	S_1	S_2	S_3
π_{S^2}	0.4	0.4	0.2

Table 3: Initial state probabilities, π

3.2 Inference in PoHMM

The main feature of PoE is its undirected graphical modelling with no direct connection among the latent variables as they only interact indirectly via observed variables. The hidden variables all the experts are rendered independent when conditioned on visible variables. So, if the inference in each of the constituent model is tractable then the inference in the product is also tractable. To generate a data point in this model, all the experts in PoE generate an observation and if they all generated the same point then it is accepted else they again generate an observation until all the experts agree to it. Therefore all the experts have some influence over the generated data. So, the inference determines the the probability that all the experts would have taken in order to generate the given observation.

3.3 Training product of experts by minimising contrastive divergence

PoE is a method of combining densities of many latent variable models. It is defined by the following formula:
To fit the model to the data, we need to maximize the likelihood of the dataset or minimise the Kullback-Liebler divergence between the real data and the fantasy data. The contrastive divergence algorithm for training the PoHMM has the following steps:

1. Calculate each model's gradient on a data point using forward backward algorithm.
2. For each model take a sample from the posterior distribution of paths through state space.
3. At each time step, multiply together the distributions and renormalize to get the reconstruction distribution at each step.
4. Draw a sample from the reconstruction distribution at

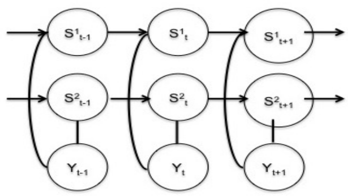

Figure 1: Product of HMMs, P = S^1 x S^2

each time step to get a reconstructed sequence. Compute each model's gradient on the new sequence.

5. Update the parameters

4. PROOF OF CONCEPTS

4.1 REDD House 2

4.2 Aim

To represent streams of energy consumption data from n[1] appliances by product of k HMMs.

4.3 Method

- **Data** The Reference Energy Disaggregation Data Set (REDD) is used in empirical analysis. The data contains power consumption from real homes, for the whole house as well as for each individual circuit in the house (labeled by the main type of appliance on that circuit). It is intended for use in developing disaggregation methods, which can predict, from only the whole-home signal, which devices are being used. The REDD data set contains two main types of home electricity data: high-frequency current/voltage waveform data of the two power mains (as well as the voltage signal for a single phase), and lower-frequency power data including the mains and individual, labeled circuits in the house. The main directory consists of several house directories, each of which contain all the power readings for a single house. Each house subdirectory consists of a labels and channels files. The labels file contains channel numbers and a text label indicating the general category of device on this channel. Each channel_i.dat file has two columns containing UTC timestamps (as integers) and power readings (recording the apparent power of the circuit) for the channel. Experiments reported here use the House 2 data from REDD. It has 11 channels where each channel corresponds to the following appliance:

  1. mains_1
  2. mains_2
  3. kitchen_1
  4. lighting
  5. stove
  6. microwave
  7. washer_dryer
  8. kitchen_2
  9. refrigerator
  10. dishwaser
  11. disposal

  The dataset has 318759 records and 2 columns. We randomly sample 300 records for our initial experiment. Time series data from two appliances are represented as product of k HMMs.

- **Time Series :** The time series data of the microwave, dryer, kitchen_2 and refrigerator are plotted below in Figures 2, 3, 4, 5.

- **Code** The implementation of the product of experts model is obtained from Iain Murray's website[2]. It implements the technique described in Geoff Hinton's paper [16].

- **Additional details** Some additional details regarding experiments:

  1. The product of HMMs model (PoHMM) minimizes "contrastive divergence" as described in the paper [16].
  2. The number of experts, k used here is 15. This is set somewhat arbitrarily and needs to be experimented on.
  3. Learning rate is $\epsilon = \frac{1}{300}$.

4.4 Experimental Setup for REDD house 2

Experiments are performed on the REDD which contains 9 appliances each containing 318759 rows of energy consumption data. Experiments are done into 4 phases, in the first phase the number of data samples are varied corresponding to which the values of KL Divergence and convergence time are noted. In the second phase, the number of experts are varied keeping the best value of the sample from the first phase fixed. In the third phase, number of iterations are varied keeping the best values from above first two phases fixed. In the fourth part, the no. of appliances to be aggregated are varied.

4.5 Results

The evaluation of how well the learning has taken place is done by using a Kullback-Leibler divergence. KL divergence of P from Q, D_{KL}(P‖Q) is the measure of information lost when Q is used to approximate P. Here, P is the real data and Q is a fantasy data. The two probability distributions in

[1] n=2

[2] http://homepages.inf.ed.ac.uk/imurray2/code/

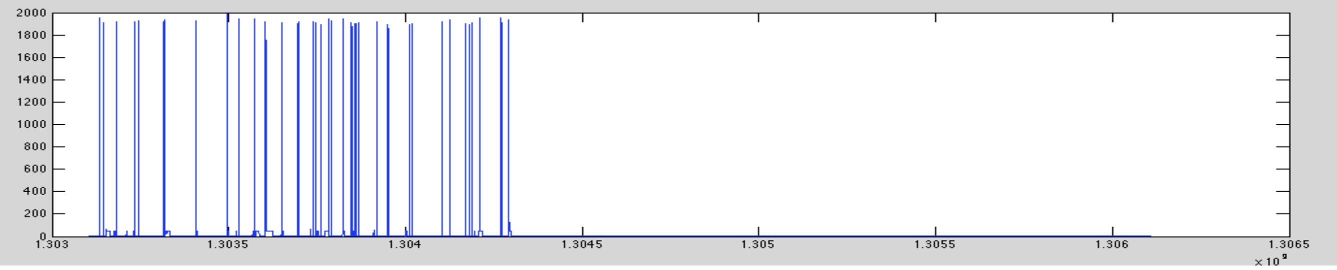

Figure 2: Microwave

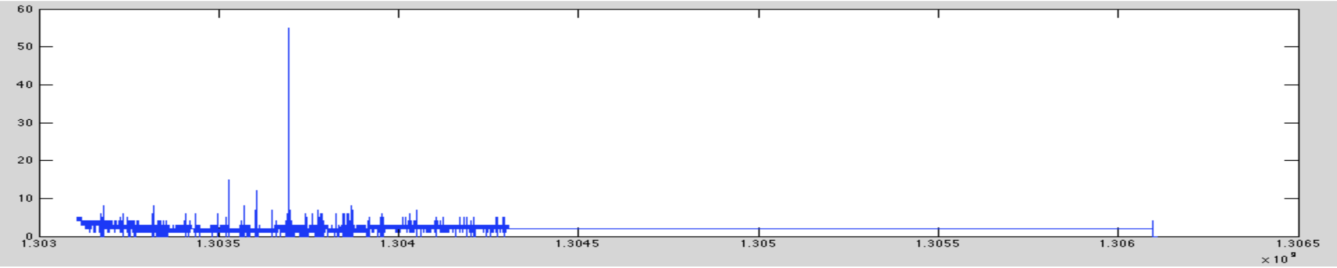

Figure 3: washer_dryer

Samples	$KLDiv$	$T(sec)$	$Iterations$
300	2.4864	186.212 ±9.087	18600
500	0.6761	106.564 ±10.046	10200
1000	1.1088	158.521 ±1.97	11200
1500	3.8829	92.896 ±8.075	5300
2000	1.8686	130.98 ±1.932	6900
2500	0.4733	215.563 ± 2.471	9900
3000	2.8204	258.213 ±1.918	11000
3500	1.2332	204.661 ±1.713	7900
4000	0.8959	292.666 ±0.619	10400
4500	1.1118	222.558 ±1.967	7200
8000	6.392	381.635 ±2.952	8100
10000	8.276	887.932 ±13.824	10500
15000	0.7201	1368.514 ±13.605	9400

Table 4: Effect of varying samples on KL div and time

Experts	$KLDiv$	$T(sec)$	$Iterations$
5	0.774	72.968 ±1.177	5200
10	1.424	117.482 ±1.966	6700
15	0.473	210.249 ±1.258	9900
20	1.56	217.739 ±10.452	9000
25	7.469	347.019 ±8.23	12100
30	2.4968	413.802 ±7.304	12900
35	1.5012	348.906 ±14.651	11300

Table 5: Effect of varying experts on KL div and time

the REDD example refer to the expert probabilities in real and fantasy data. The learned parameters from the training are fitted to the fantasy data to measure the information lost when fantasy data is used to approximate real data.

Threshold	$KLDiv$	$T(sec)$	$Iterations$
.1	0.473	210.6 ±1.493	9900
.05	0.443	240.607±2.436	10900
.01	0.454	431.536 ±14.509	18000
.005	0.509	1167.243 ±43.412	49800

Table 6: Effect of varying min threshold on KL div and time

Appliances	$KLDiv$	$T(sec)$	$Iterations$
3	5.559	233.664 ±0.579	10700
4	0.188	465.634 ±5.275	19900
5	.432	338.416 ±3.988	13400
6	8.736	606.062 ±7.534	28100
7	5.054	411.457 ±10.051	17300
8	0.436	260.544 ±cc27.862	10700
9	0.15	474.579 ±14.619	20600

Table 7: Effect of varying appliances on KL div and time

5. PROOF OF CONCEPT ON FACULTY HOUSING DATA

5.1 Aim

To represent streams of energy consumption data from all the floors of faculty housing as a product of k HMMs.

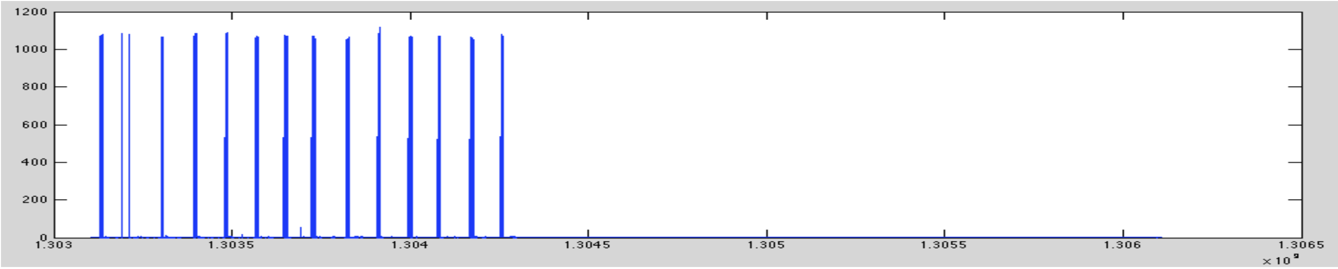

Figure 4: Kitchen_2

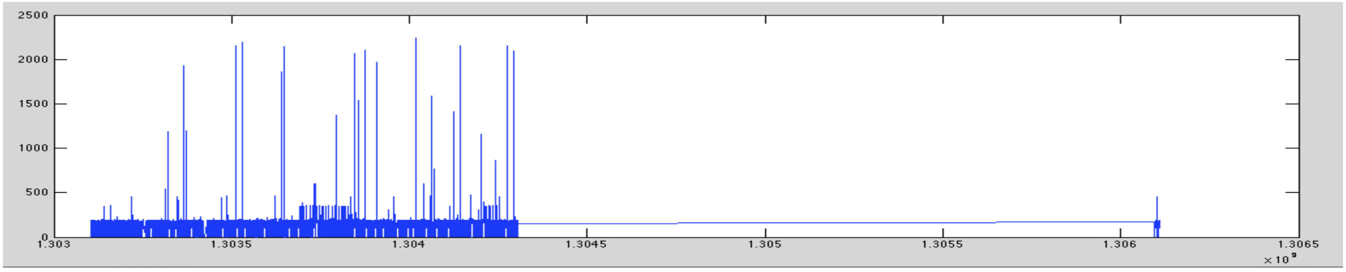

Figure 5: refrigerator

5.2 Method

- **Data** This data represents the energy consumed by the IIIT Delhi faculty housing building. As a part of research, a team from IIIT Delhi has installed various temperature, light and motion sensors to perform real world studies and to analyse user preferences for energy conservation. For our analysis, we selected one month's historical data ranging from 01-01-2014, 00:01 hours to 31-01-2014, 23:59 hours. The two smart meters installed captures the data from all the floors. The first meter gives out readings from floors 0 to 5 and the second meter gives out readings from floors 6 to 11. The dataset includes timestamp and power consumed in watts and 84133 records. Time series data from two streams are modelled as a product of k HMMs. We also have the total power consumed by the faculty housing building which would serve as the ground truth to compare product of k HMMs with. The data is obtained from the website whose screenshot is shown in Figure 6

- **Code** It implements the technique described in Geoff Hinton's paper [16].

- **Time Series :** The time series data of the energy consumption of floor 0 to 5, floor 6 to 11 and total power are plotted below in Figures 7, 8, 9.

- **Additional Details**

5.3 Experimental Setup

Each of the data stream is modelled as a HMM individually. There are three streams of data, the first stream D_1 corresponds to the data from 0-5th floor, D_2 corresponds to 6-11th floor and D_3 represents the total power from the faculty housing which is represented by a fixed test set, T. Firstly, the stream D_1 is used to train the model such that the contrastive divergence is minimized. The parameters (mixing component of each unigauss, means of gaussian bits, log precisions of axis-aligned gaussian bits) that are learnt during the training are provided to the test set T in order to obtain the conditional probability of the gaussians given the data D_1 represented as $pgauss_1$. Similarly, the second stream of data, D_2 collected from floor 6-11, is used to learn the parameters of the model during the training phase which are then again provided to the test set T to obtain conditional probability of the gaussians given the data D_2 as $pgauss_2$. Finally the data D_3 is used to learn the model and parameters which are then applied to the test set T to obtain the gaussian probability as $pgauss_3$. Now, as we know that the total power consumption of the building should be approximately equal to the product of HMMs, which is the product of $pgauss_1$ and $pgauss_2$. If we can show that the value of $pgauss_3$ is as close as possible to the product of $pgauss_1$ and $pgauss_2$.
The experiments performed in table 8 , shows the effect of varying samples on KL Divergence, convergence time and iterations keeping minimum threshold constant at 7.
The other experiment performed in table 9 shows that effect of varying experts on KL Divergence and convergence time.

5.4 Results

Table 8 shows that the error was minimum when the sample size was 300. With respect to the number of experts, the error was minimum when there were 5 experts as shown in table 9.

6. CONCLUSIONS

7. ACKNOWLEDGMENTS

8. REFERENCES

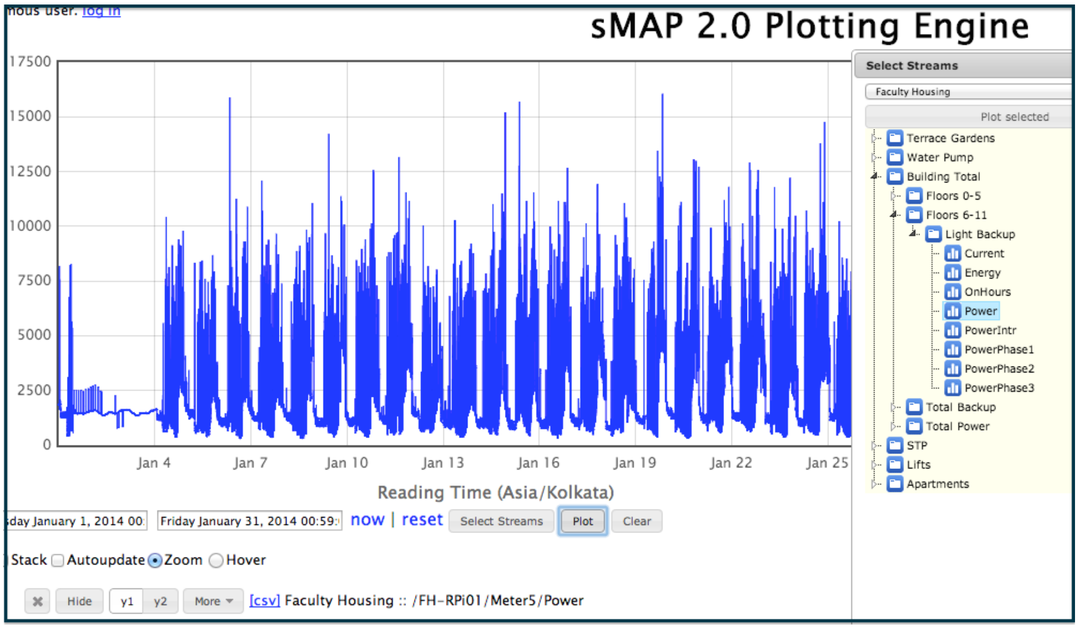

Figure 6: Screen shot of the webpage

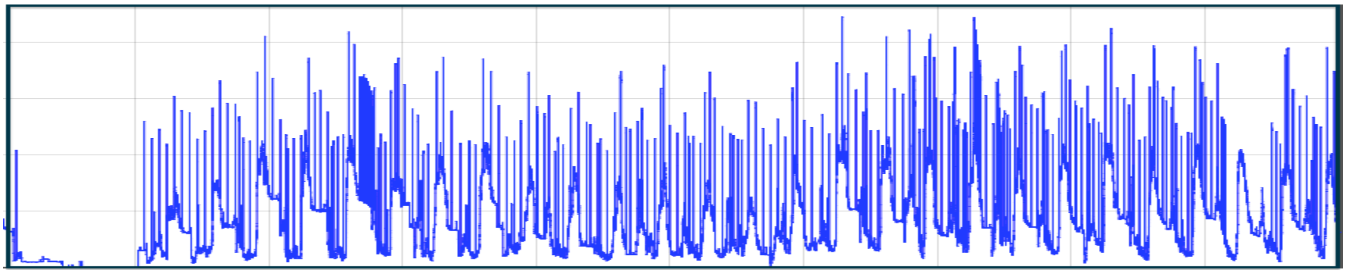

Figure 7: Stream 1: Power consumption of floors 0-5

Samples	$KLDiv$	$T(sec)$	$Iterations$
100	2.6219e-05	257	45100
300	1.9753e-05	222	43200
500	5.5493e-05	260	44800
700	3.2847e-05	249	44000
900	3.9486e-04	221	42600
1100	4.9274e-04	317	44700
1300	3.0425e-04	276	43100
1500	3.1128e-04	303	44400
2000	1.9192e-04	306	44400
2500	1.7122e-04	370	44100
3000	1.4686e-04	331	43300
3500	1.2663e-04	370	43200
4000	1.0793e-04	403	43200

Table 8: Effect of varying samples on KL div

Experts	$KLDiv(e-05)$	$T(sec)$
3	1.9780	229
4	3.5897	217
5	1.9753	228
6	4.3488	238
7	4.9111	245
8	5.6564	241
9	5.4290	258
10	5.5163	267
12	4.4504	262
14	6.9006	296
16	6.8666	300
18	6.2872	313
20	5.3842	267
25	5.8970	326
30	5.9962	327
35	5.2716	346
40	5.0955	320

Table 9: Effect of varying experts on KL div and time

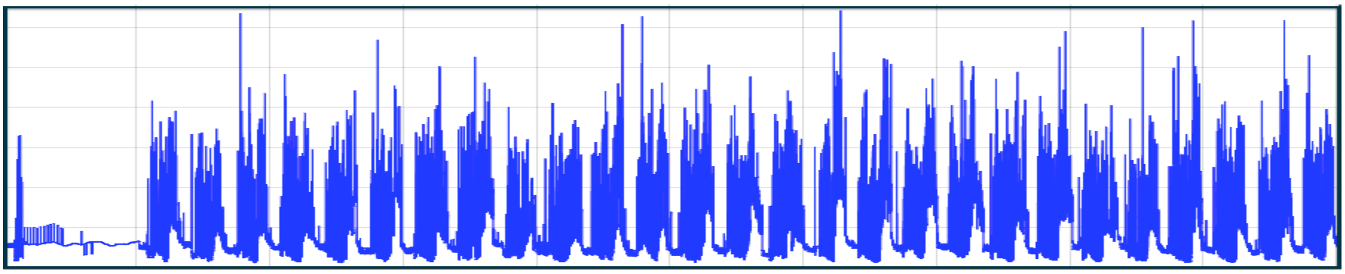

Figure 8: Stream 2: Power consumption of floors 6-11

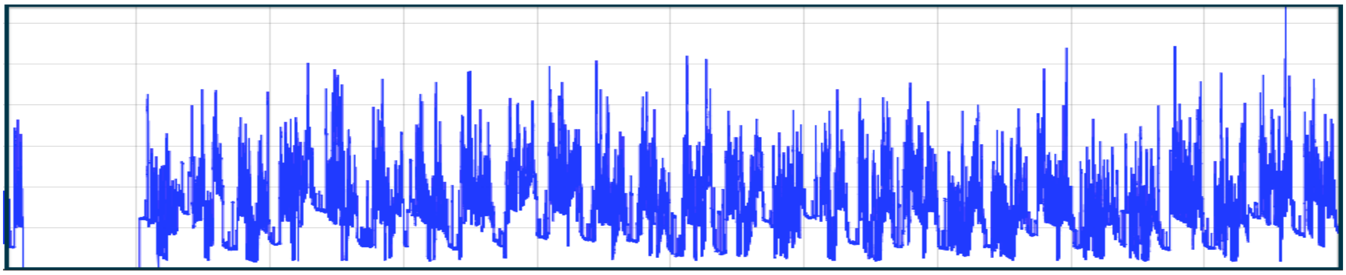

Figure 9: Total Power of the building

[1] A. Albert and R. Rajagopal. Smart meter driven segmentation: What your consumption says about you. *Power Systems, IEEE Transactions on*, 28(4):4019–4030, Nov 2013.

[2] A. Bakirtzis, V. Petridis, S. Kiartzis, and M. Alexiadis. A neural network short term load forecasting model for the greek power system. *Power Systems, IEEE Transactions on*, 11(2):858–863, May 1996.

[3] D. Bassi and O. Olivares. Medium term electric load forecasting using tlfn neural networks. *International Journal of Computers, Communications & Control*, 1(2):23–32, 2006.

[4] A. Brown and G. Hinton. Proceedings of artificial intelligence and statistics 2001. In *Products of Hidden Markov Models*, number GCNU TR 2000-008, 2001.

[5] A. D. Brown. Product model for sequences, 2001.

[6] H. Chen, C. Canizares, and A. Singh. Ann-based short-term load forecasting in electricity markets. In *Power Engineering Society Winter Meeting, 2001. IEEE*, volume 2, pages 411–415 vol.2, 2001.

[7] G. Chicco, R. Napoli, and F. Piglione. Comparisons among clustering techniques for electricity customer classification. *Power Systems, IEEE Transactions on*, 21(2):933–940, May 2006.

[8] T. Chow and C. Leung. Nonlinear autoregressive integrated neural network model for short-term load forecasting. *Generation, Transmission and Distribution, IEE Proceedings-*, 143(5):500–506, Sep 1996.

[9] H. Daneshi, M. Shahidehpour, and A. Choobbari. Long-term load forecasting in electricity market. In *Electro/Information Technology, 2008. EIT 2008. IEEE International Conference on*, pages 395–400, May 2008.

[10] E. Fabre, S. Haar, A. Benveniste, and A. Benveniste. Hidden markov models for distributed and concurrent systems, 2000.

[11] M. Falvo, R. Lamedica, S. Pierazzo, and A. Prudenzi. A knowledge based system for medium term load forecasting. In *Transmission and Distribution Conference and Exhibition, 2005/2006 IEEE PES*, pages 1291–1295, May 2006.

[12] M. D. Felice and X. Yao. Short-term load forecasting with neural network ensembles: A comparative study [application notes]. *IEEE Comp. Int. Mag.*, 6(3):47–56, 2011.

[13] Z. Ghahramani. Hidden markov models. pages 9–42, 2002.

[14] Z. Ghahramani and M. I. Jordan. Factorial hidden markov models. *Mach. Learn.*, 29(2-3):245–273, Nov. 1997.

[15] W. R. Heinzelman, A. Ch, and H. Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. pages 3005–3014, 2000.

[16] G. E. Hinton. Training products of experts by minimizing contrastive divergence. Technical Report GCNU TR 2000-004, Gatsby Computational Neuroscience Unit, University College London, 2000.

[17] H. Kim, M. Marwah, M. F. Arlitt, G. Lyon, and J. Han. Unsupervised disaggregation of low frequency power measurements. In *SIAM*, pages 747–758. SIAM / Omnipress, 2011.

[18] J. Z. Kolter and J. Ferreira. A large-scale study on predicting and contextualizing building energy usage. In *AAAI*. AAAI Press, 2011.

[19] J. Z. Kolter and T. Jaakkola. Approximate inference in additive factorial hmms with application to energy disaggregation. In *AISTATS*, volume 22 of *JMLR Proceedings*, pages 1472–1482. JMLR.org, 2012.

[20] F. Martinez Alvarez, A. Troncoso, J. Riquelme, and J. Aguilar Ruiz. Energy time series forecasting based on pattern sequence similarity. *Knowledge and Data*

Engineering, IEEE Transactions on, 23(8):1230–1243, Aug 2011.

[21] C. McKerracher and J. Torriti. Energy consumption feedback in perspective: integrating australian data to meta-analyses on in-home displays. *Energy Efficiency*, 6(2):387–405, 2013.

[22] P. Mirowski, S. Chen, T. Kam Ho, and C.-N. Yu. Demand forecasting in smart grids. *Bell Labs Technical Journal*, 18(4):135–158, 2014.

[23] I. A. Samuel, A. A. Awelewa, et al. Medium-term load forecasting of covenant university using the regression analysis methods. *Journal of Energy Technologies and Policy*, 4(4):10–16, 2014.

[24] W. Shen, V. Babushkin, Z. Aung, and W. L. Woon. An ensemble model for day-ahead electricity demand time series forecasting. In *Proceedings of the Fourth International Conference on Future Energy Systems*, e-Energy '13, pages 51–62, New York, NY, USA, 2013. ACM.

[25] D. I. Stern. The role of energy in economic growth. *Annals of the New York Academy of Sciences*, 1219(1):26–51, 2011.

[26] G. Taban and A. A. C?rdenas. Data aggregation as a method of protecting privacy in smart grid networks. *Power Systems, IEEE Transactions on*, March 2012.

[27] G. W. Taylor and G. E. Hinton. Products of hidden markov models: It takes n >1 to tango. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, pages 522–529, Arlington, Virginia, United States, 2009. AUAI Press.

[28] T. K. Wijaya, J. Eberle, and K. Aberer. Symbolic representation of smart meter data. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, EDBT '13, pages 242–248, New York, NY, USA, 2013. ACM.

[29] T. K. Wijaya, T. Ganu, D. Chakraborty, K. Aberer, and D. P. Seetharam. Consumer segmentation and knowledge extraction from smart meter and survey data. In *SIAM International Conference on Data Mining (SDM14)*, number EPFL-CONF-196276, pages 226–234, 2014.

[30] Z. Zhang and S. Ye. Long term load forecasting and recommendations for china based on support vector regression. In *Information Management, Innovation Management and Industrial Engineering (ICIII), 2011 International Conference on*, volume 3, pages 597–602, Nov 2011.