Identification of the Potential of Residential Demand Response Using Artificial Neural Networks

Ashwin Shirsat
Department of Electrical and Computer Engineering
North Carolina State University
Raleigh, NC 27695, USA
ashirsa@ncsu.edu

Wenyuan Tang
Department of Electrical and Computer Engineering
North Carolina State University
Raleigh, NC 27695, USA
wtang8@ncsu.edu

***Abstract*—The uncertain and volatile nature of renewable energy tends to cause imbalance in the supply and demand of electricity. Residential demand response has emerged as a methodology to alleviate such imbalances in the electricity grid. By utilizing the flexibility of the electricity demand, residential demand response programs offer monetary incentives to the customers to bring about temporary reductions in energy consumption when the grid faces supply shortage or is overloaded. Utilities are interested in enrolling residential customers in their demand response programs in order to maintain a healthy demand-supply balance. Selecting the right set of customers for this is a demanding task, and what makes it even more challenging is the lack of use of the temporal consumption data in the existing customer recruiting methodologies. In this paper, by utilizing the granular smart meter data, we develop a methodology that estimates a customer's potential reduction in consumption based on his/her sensitivity to external temperature, which is then used to train a mixture density recurrent neural network. The network outputs the probability distribution of the demand reduction of the customer. The results can be utilized to accurately select customers, which can then ensure the reliable performance of the demand response program.**

I. Introduction

The smart grid is conceived as a modern electric grid that can deliver energy in a controlled and an optimized manner from the points of generation to the points of consumption. This makes the consumers an integral part of the grid since they can alter their consumption patterns according to the information and incentives provided to them. Hence, demand side management (DSM) plays a vital role in the smooth operation of the smart grid. DSM programs are enabled by the integration of the information and communication technology, which establishes a bidirectional communication link between the generators and consumers [1].

One of the prime DSM programs is the demand response (DR) program. The U.S. Department of Energy defines DR as a program established to motivate changes in the electricity use by the end-use customers in response to changes in the electricity price, or the incentive payments designed to promote lower electricity consumption at times of high market prices or when the grid reliability is jeopardized [2]. The underlying idea is to make DR attractive to the customers for them to manage their consumption pattern in a manner that would not only benefit them, but also the power grid [3].

Since customers are the key participants in a DR program, their effective recruitment is an essential prerequisite for its success. The existing techniques to target and enroll potential customers utilize predictive market models that use socio-economic factors such as family income, household size, age, education, presence of children, education, enrolment in other energy-efficiency programs, and historical energy bill costs [4]. Although the granular consumption data is available due to the widespread installation of smart meters, it has not been effectively used to target potential customers.

In the United States, the California Public Utilities Commission launched a demand response auction mechanism (DRAM) program for the state of California in July 2015 [5]. DRAM requires utilities to acquire a minimum monthly amount of reduction capacity from DR aggregators. The real-time market decides the electricity prices by matching the demand-supply curves subject to the acquired day-ahead reduction capacity. Once a utility's bid is cleared, it asks its DR aggregator to incentivize its customers to reduce their consumption. To effectively realize the bid and make a profit, the DR aggregator needs to accurately estimate the achievable reduction. The reduction in consumption highly depends on the set of customers that the DR aggregator chooses. The customer selection directly affects the reliability as each one will provide different level of uncertainty on the load reduction they can offer during a DR event [6]. The DR aggregator can improvise his bidding strategy by modeling the customers' consumption behavior and then selecting those customers that show a high potential for reduction during DR hours.

There is a significant amount of literature describing various aspects of DR such as its definition [7], design of DR schemes [8], pricing and optimization algorithms for DR implementation [9], baseline calculation [10], and other areas. On topics related to targeting customers for their participation in DR schemes, the authors in [6] estimate the reduction response based on the customer's thermal sensitivity by using weather and consumption data. After obtaining the customer response, they solve a stochastic knapsack problem to choose the optimal set of customers. However, their methodology does not effectively account for the uncertainty in the customer behavior. In [11], the authors have used machine learning methods for the prediction of residential energy consumption along

978-1-7281-0407-2/19/$31.00

with non-parametric hypothesis test to estimate customers' reduction during DR hours. Using actual DR data, the authors conclude that the customers who demonstrate highly variable consumption patterns have a high probability of reducing their consumption as opposed to those having a regular consumption pattern. However, their research lacks the accurate quantification of the likelihood with which the customers will reduce their consumption. In this paper, we develop a methodology to effectively quantify the load reduction potential of residential customers based on their temporal consumption by combining data analytics and neural networks, that can be used by a DR aggregator to select customers on a daily basis. The contribution of this paper can be summarized as follows:

1) Analyzing and quantifying a customer's electricity consumption sensitivity to the external temperature;
2) Estimating the potential consumption reduction of a customer based on their temperature sensitivity;
3) Training a mixture density recurrent neural network (MD-RNN) to estimate the probability distribution of the potential reduction using the consumption data and spatiotemporal features.

The rest of the paper is organized as follows. Section II describes the methodology used to model the customers' thermal response, provides preliminary background on recurrent neural networks (RNNs) and mixture density networks (MDNs), and describes the MD-RNN model. Experimental results of our proposed methodology are presented in Section III. Section IV concludes the paper.

II. Methodology

This paper investigates on how to effectively obtain the potential of a customer to respond to DR signals in a thermal DR program. The thermal DR program belongs to the category of "Global Temperature Adjustment" and is an effective strategy for systems involving HVAC units [12]. We adopt a thermal response model from [6] and [13] to obtain the consumption sensitivity of the customer to changes in the external temperature. This model helps us estimate the possible reduction in consumption during a DR event. Using the consumption, local weather, and temporal information as the features along with the estimated energy savings as the labels, we train a long short-term memory (LSTM) based RNN model. Since the estimated reduction in consumption of a customer will not necessarily be equal to the actual reduction, the output of the RNN is not very intuitive and cannot be relied upon completely. Hence, we pass the output of the RNN through a MDN which outputs the probability distribution of the potential consumption reduction of the customer. Using the probability distribution, an appropriate set of customers for DR can be chosen in a probabilistic manner.

A. Customer Thermal Response Modeling

We begin by estimating the customer response, i.e., the possible reduction in cooling/heating power consumption that can be obtained by increasing/decreasing the HVAC set point. To do so, we first begin by analyzing the parameters that quantify the HVAC consumption sensitivity to the external temperature. A linear model is developed to estimate the sensitivity coefficients, using which the response r_k is calculated for each customer k [13]. Since the reduction is a random variable, the accuracy of the linear model does not affect the performance of the proposed methodology [6]. HVAC system consumption has the highest contribution towards the total residential load. A piecewise linear relationship with two break points is observed between the energy consumption and the external temperature [6]. Heating and cooling power consumption increases linearly with temperature. The consumption of the customer k at time t on day d, modelled as a linear function of the external temperature and the break point, can be written as follows:

$$l_k(t,d) = a_k(t)(T_k^o(t,d) - T_k^s)_+ + b_k(t)(T_k^s - T_k^o(t,d))_+ \\ + c_k(t) + e_k(t). \quad (1)$$

Herein, $T_k^o(t,d)$ is the external temperature, and T_k^s is the break point temperature, which can be regarded as the HVAC set point. The coefficient $a_k(t)$ is the cooling sensitivity, $b_k(t)$ is the heating sensitivity, $c_k(t)$ is the base load, and $e_k(t)$ is the error. $a_k(t)$ and $b_k(t)$ are approximately zero during the seasons of winter and summer, respectively. The base load corresponds to the consumption of the house excluding the consumption of the HVAC system. The error term is assumed to follow a normal distribution. It is to be noted that the coefficients take values that vary through time, but are fixed for all the days of a particular season. Since T_k^s can be considered equivalent to the HVAC set point, there are limited values which it can take and they have to be integers with an increment of 1 °F. We assume T_k^s to take an integral value in the range 70–85 °F based on the normal operation set points for a standard HVAC system, which also simplifies the computation process. The regression based model learning is performed in two stages for every customer k. In the first stage, the model parameters $\{a_k(t),\ b_k(t),\ c_k(t)\}$ are learnt for every possible value of T_k^s and the residual sum of squares (RSS) is computed. In the second stage, the model parameters corresponding to the minimum RSS value are selected. The total computation complexity for the training process equals to that of a simple linear regression complexity times the total number of break point values.

Once the model parameters are estimated, we utilize them to formulate the response model. The reduction in consumption is obtained by increasing/decreasing the HVAC set point by ΔT °F. Let us assume a DR event on a hot summer day, i.e., $T_k^o(t,d) \geq T_k^s + \Delta T$. The response can be then calculated as $r_k(t) = a_k(t)\Delta T$ kWh, where ΔT is the ordered increase in set point. Similarly, on a cold day, the response will be $r_k(t) = b_k(t)\Delta T$, where ΔT is the ordered decrease in the set point. A positive value of $r_k(t)$ indicates reduction in consumption and vice versa.

B. Recurrent Neural Network (RNN)

Neural networks (NNs) that contain feedback connections from the neurons back to themselves are called RNNs, in which the behavior of the neurons is time dependent. Due to this, RNNs have been theoretically proven to be able to approximate the mapping amongst time series to an arbitrary accuracy under certain conditions. The prediction made by an RNN at time slot $t-1$ impacts the prediction it will make at time t. Hence, for making predictions for the future, RNNs have two input sources, the present and the recent past. LSTMs are a special kind of RNNs that are capable of learning the long-term dependencies and have recently gained popularity in many applications. LSTMs allow for a better control over gradient flow by overcoming the vanishing and exploding gradient problem by introducing new gates. At any given time t, let c_t, h_t, and x_t be the LSTM cell's internal state, hidden state, and input, respectively. c_t interacts with h_t and x_t to decide which elements of the internal state vector need to be updated, erased, or maintained based on the outputs of time slot $t-1$. A detailed description on LSTM cells can be found in [14].

C. Mixture Density Network (MDN)

Minimizing cross-entropy error function or sum-of-squares produces network outputs that can approximate the conditional averages of the target data on the basis of the given input vector. These averages can be regarded as optimal for a classification problem since they represent the posterior probabilities of a class membership [15]. If the nature of the problem revolves around the prediction of continuous variables, the conditional averages provide a very limited information about the target variables. This is predominantly observed when the mapping to be learned is multi-valued, since averaging several correct target values will not necessarily produce a correct value. This problem can be overcome if we model the conditional probability distribution of the target data conditioned on the input vector. An MDN, the combination of a conventional neural network with a mixture density model, has the capability of representing arbitrary conditional probability distributions in the same way a conventional neural network represents arbitrary functions. The prediction of the consumption reduction potential of a customer can be seen as a multi-valued problem, since on different days, with different consumption profile and weather, the reduction in consumption could be the same. In short, a same consumption reduction value can be taken by different feature sets. This makes the application of MDN for this problem an appropriate one.

We will briefly explain the simplified equations for an MDN with single dimensional output, which in our case will be the customer's consumption reduction. The analysis is carried out on a per customer basis, hence the subscript k has been omitted for simplicity. Let $\mathbf{X}$ and $\mathbf{r}$ be sequences of length T, where $\mathbf{x} \in \mathbf{X}$ is the K dimensional input vector, $\mathbf{r}$ is the one dimensional target vector, and T is the total number of hourly time slots within the horizon. We assume r^t is independent for all $t \in T$ conditioned on $\mathbf{x}^t$:

$$p(\mathbf{r}|\mathbf{X}) = \prod_{t=1}^{T} p(r^t|\mathbf{x}^t). \tag{2}$$

The probability density of the target data is then represented as a linear combination of kernel functions:

$$p(r^t|\mathbf{x}^t) = \sum_{m=1}^{M} p(m|\mathbf{x}^t)p(r^t|\mathbf{x}^t, m), \tag{3}$$

where M is the total number of components in the mixture, and $p(m|\mathbf{x}^t)$ are the mixing coefficients which will be denoted by $\gamma_m(\mathbf{x}^t)$. Mixing coefficients are the probabilities of choosing the mixture component $m \in M$ given $\mathbf{x}^t$. $p(r^t|\mathbf{x}^t, m)$ denotes the conditional density of the target value r^t for the m^{th} kernel. In this paper, we restrict our analysis to kernel functions that are Gaussian in nature:

$$p(r^t|\mathbf{x}^t, m) = \frac{1}{\sqrt{2\pi\sigma_m(\mathbf{x}^t)^2}} \exp\left(-\frac{(r^t - \mu_m(\mathbf{x}^t))^2}{2\sigma_m(\mathbf{x}^t)^2}\right), \tag{4}$$

where $\mu_m(\mathbf{x}^t)$ and $\sigma_m(\mathbf{x}^t)$ denote the mean and standard deviation corresponding to the m^{th} distribution in the mixture. The parameters of the mixture model, namely the mixing coefficients $\gamma_m(\mathbf{x}^t)$, the means $\mu_m(\mathbf{x}^t)$, and the standard deviations $\sigma_m(\mathbf{x}^t)$, are taken to be continuous functions of $\mathbf{x}^t$. This is achieved by modeling them using the outputs of a conventional neural network. By choosing a mixture model with the appropriate number of kernel functions and a neural network with sufficient number of hidden units, the MDN can closely approximate any conditional probability $p(r^t|\mathbf{x}^t)$. In this work, we considered a multi-layer perceptron with 24 hidden layers of 12 activation units, and a linear output layer. For a one-dimensional target variable, the total number of outputs of the MDN is given by $3M$, which, in our case, equals 36. The mixing coefficients $\gamma_m(\mathbf{x}^t)$ must satisfy the following constraint:

$$\sum_{m=1}^{M} \gamma_m(\mathbf{x}^t) = 1, \qquad 0 \leq \gamma_m(\mathbf{x}^t) \leq 1, \tag{5}$$

which can be enforced by using a softmax activation function for the output layer. The means $\mu_m(\mathbf{x}^t)$ represent the location parameters and have no constraints on the values they can take:

$$\mu_m(\mathbf{x}^t) = z_m^{\mu}, \tag{6}$$

where z_m^{μ} represent the corresponding network outputs. The standard deviations $\sigma_m(\mathbf{x}^t)$ represent the scale parameters and can be expressed as exponentials of the corresponding network outputs:

$$\sigma_m(\mathbf{x}^t) = \exp(z_m^{\sigma}). \tag{7}$$

We then construct an error function to facilitate the MDN training by taking the negative logarithm of the likelihood:

$$E = -\sum_{t=1}^{T} \ln\left(\sum_{m=1}^{M} \gamma_m(\mathbf{x}^t)p(\mathbf{r}^t|\mathbf{x}^t, m)\right). \tag{8}$$

The objective of the MDN is to minimize this error function. Additional details on minimizing the error function in (8) can be found in [15].

D. Mixture Density Recurrent Neural Network (MD-RNN)

An MDN's output parameterizes a mixture of Gaussian distributions for the target variable, which gives more insight when the target variable is continuous in nature. However, a simple MDN does not incorporate the ability to learn and make time series predictions. The popular LSTM architecture works best for analyzing time series data. Since our DR application requires the analysis of time series data and prediction of a continuous variable, we build an MD-RNN, in which the outputs of an RNN are fed to the MDN to obtain the distribution of the target variables. In the MD-RNN, the recurrent part allows the modeling of time series data, while the Gaussian mixture part allows the predictions to be creative [16]. In short, the MD-RNN model combines a conventional LSTM based RNN with a loss function that maximizes the conditional probability $p(\mathbf{r}|\mathbf{x})$ [17]. Algorithm 1 gives the training procedure for the MD-RNN.

Algorithm 1 Algorithm for training the MD-RNN

Input: normalized input features ($\mathbf{X}$), original labels (y), normalized labels ($\bar{y}$), total number of MDN mixtures (M)
Output: mixture weights (γ_m), mixture means (μ_m), mixture standard deviations (σ_m)
Initialization :
Convert $\mathbf{X}$ into a 3D format, $\mathbf{X}_{3D}$, of the dimension $T \times n \times K$, where n is the predetermined number of time steps
Associate ($\bar{y}$) with each sample in (X_{3D})
for k epochs **do**
    Train the LSTM network using X_{3D} and ($\bar{y}$) for 1 epoch
    Extract the output ($\bar{y}_{pred}$) of the network and denormalize it (y_{pred})
    Train the MDN using the input feature y_{pred}, the label y and the loss function (8)
    Reset the LSTM states
end for

III. Experimental Results

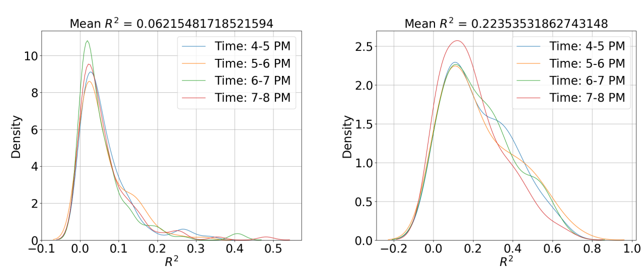
In this section, we verify the proposed methodology for analyzing the reduction potential of a customer. We first describe the data used for training and testing. We then demonstrate and analyze the results of the MD-RNN model. All the simulations were performed using Python with Keras, an open-source deep learning library.

A. Data Description

The residential consumption data from the UMass Smart Apartment dataset [18] is used for testing our proposed approach. The dataset comprises of residential consumption data collected from smart meters installed in 114 apartments. The data spans over two years ranging from January 2015 to December 2016. For the first year, the daily consumption data is provided at an interval of 15 minutes, while for the second year, it is provided at an interval of 1 minute. The dataset also contains weather information at an hourly interval for both the years. To match the interval spacing of the weather data with that of the consumption data, we process the consumption data to take values on an hourly basis by averaging the 15 minute interval values for each hour. The apartments are located in Western Massachusetts, which sees cold winters with temperatures around freezing and relatively cooler summers. We further process this raw data by classifying it into seasons and removing the data points for weekends and holidays.

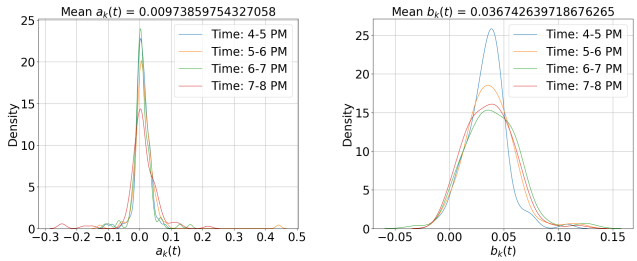
B. Customer Response Model Fitting

We fit the regression model (1) using the consumption data, and obtain values for the break points and the heating/cooling sensitivities. The load data indicates that the peak consumption hours for the 114 residential apartments occur during 2PM–10PM. These are the hours a utility will mostly require to initiate DR. We fit the response model for these apartments for two seasons, i.e., summer (June–September) and winter (December–March). We present the results for a subset of 4 hours (4PM–8PM) of the total peak consumption period. We begin by analyzing the quality of the fitted model using the R^2 (coefficient of determination) regression score. Fig. 1 shows the R^2 values for both the seasons for all the customers. We observe that the variance is higher during the winter season as compared to the summer season. The possible explanation for this could be that during the winter season, a variable electricity consumption is observed due to frequent use of electricity for space heating, while during the relatively cooler summers, the air conditioning is used less often. Fig. 2 shows the distribution of $a_k(t)$ and $b_k(t)$ for all the customers for the summer and winter seasons, respectively. It can be observed that during summer, the cooling sensitivity of customers is narrowly centered around 0, while the heating sensitivity during winter is largely on the positive side. This is suggestive of the fact that the potential reduction capability via a thermal DR program is higher during winter. Fig. 3 shows the distribution of $a_k(t)/\sigma(a_k(t))$ and $b_k(t)/\sigma(b_k(t))$ for summer and winter, respectively. The value of this ratio is higher for the winter season, which translates to the realization of relatively stable energy savings during DR events. By analyzing all the different metrics, it is observed that the DR potential is higher during the winter season for this particular data set as well as for areas that face a cooler climate. On the contrary, a high DR potential can be realized during summer season if the residential area faces a hot climate as shown in [6].



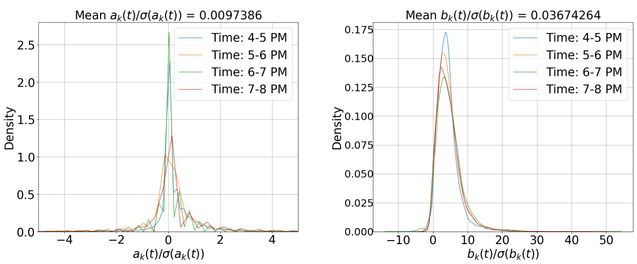
(a) (b)

Fig. 1. R^2 density plots for the year 2015. (a) Summer and (b) Winter.



(a) (b)

Fig. 2. Temperature sensitivity density plots for the year 2015. (a) Summer and (b) Winter.



(a) (b)

Fig. 3. Temperature sensitivity variation density plots for the year 2015. (a) Summer and (b) Winter.

C. *Load Reduction Potential Prediction using MD-RNN*

We first begin by describing the features and labels used for training the MD-RNN model. For every time slot, a vector of 32 features is generated. The first two features include the residential consumption and the external temperature. The next 29 features represent the hour of the day (1–24) and the day of the week (1–5) in an one hot encoded fashion. The final feature corresponds to the break point temperature value. The heating and cooling sensitivities are obtained using the methodology described in Section II.A. The reduction in consumption is then calculated for a change in the HVAC temperature set point (ΔT) of 3 °F. For summer season, the temperature set point will be raised by 3 °F, and reduced by 3 °F for the winter season. The final reduction value, obtained by multiplying ΔT with the heating or cooling sensitivity value, forms the training label. From the observed data, the load starts to rise around 2PM, with the peak hours extending from 4PM–10PM. Hence, for the DR hours between 2PM–10PM, we use the labels as obtained above, while the labels for non-DR hours are set to zero. We also add normally distributed random error to ensure that the simulation is as realistic as possible. The training data corresponds to the summer and winter seasons for the year 2015. The respective season data from the year 2016 is used for making the predictions. Since the purpose of this model is to make day ahead predictions on the customers' potential, the testing data would ideally be the forecasted consumption and temperature values. The testing labels are generated using the sensitivities obtained from fitting the regression model of Section II.A using the data of the year 2015 for a change in temperature set point (ΔT) of 3 °F. The sensitivities and break point values will be constantly updated along with the addition of new data to the training set as and when days go by. In this way, the model will be able to continuously learn on the fly. To train the LSTM model, the data needs to be in a three-dimensional format. The first dimension (T) is the number of training samples, which is equal to the total number of time slots. The second dimension is the time steps (n) in each training sample. The time steps ensure that each training sample for the LSTM model contains information pertaining to the current time slot and $n-1$ time slots of the recent past. We choose 4 time steps per sample, since on average, a single DR period lasts for about 4 hours. The third dimension is the total number of features (K) for every training sample. To prevent overfitting, we insert dropout for every layer in the LSTM model. For the MDN, we choose 24 hidden layers and output a mixture consisting of 12 distributions. The model is then trained using the method shown in Algorithm 1 for 250 epochs.

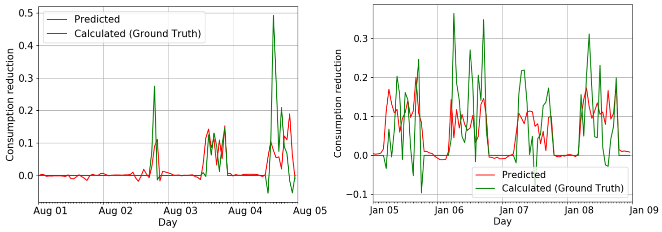
TABLE I
MD-RNN MODEL ERROR

Apt. No.	**Summer**		**Winter**	
	Train	Test	Train	Test
Apt. 1	0.0173	0.0186	0.0112	0.0128
Apt. 2	**0.0096**	**0.0109**	**0.0085**	**0.0113**
Apt. 3	0.0091	0.0131	0.0098	0.0135
Apt. 4	0.0176	0.0228	0.0215	0.0289
Apt. 5	0.0086	0.0119	0.0115	0.01449

The training and testing error of the model, when analyzed on 5 different apartments, is shown in Table I. From the error values, we can conclude that the models have been trained properly and make near accurate predictions. The detailed plots of the model performance for Apartment 2 are shown in Figs. 4–6. From Fig. 4, we see that the LSTM network is able to appropriately follow the trend and predict the customer consumption reduction values close to the calculated ground truth values. Figs. 5 and 6 show the probability density for the potential reduction in consumption during the DR hours from 2PM–10PM. From these plots, we see that the expected mean value of the mixtures deviates slightly from the ground truth value. Moreover, we observe that the probability density plot for the summer season is very narrow. This is due to the fact that the customers utilize the cooling systems to a lesser extent due to the relatively cooler summers. On the contrary, a wider distribution plot for the winter season is observed due to the heavy reliance on the heating systems. It is evident from these plots that the winter season has a higher potential for consumption reduction as compared to the summer season. These results adhere to those we would expect from the response model described in Section II.A and III.B.

IV. Conclusion

In this paper, we use linear regression to obtain the thermal sensitivity of the customer from the weather and the consumption data. Using this sensitivity along with the concept of MD-RNN, we develop a methodology for learning the consumption reduction capacity of different customers and generating mixture distributions to better quantify the reduction potential. The



(a) (b)

Fig. 4. LSTM model results for Apt. 2. (a) Summer (Aug. 1–4, 2016) and (b) Winter (Jan. 5–8, 2016).

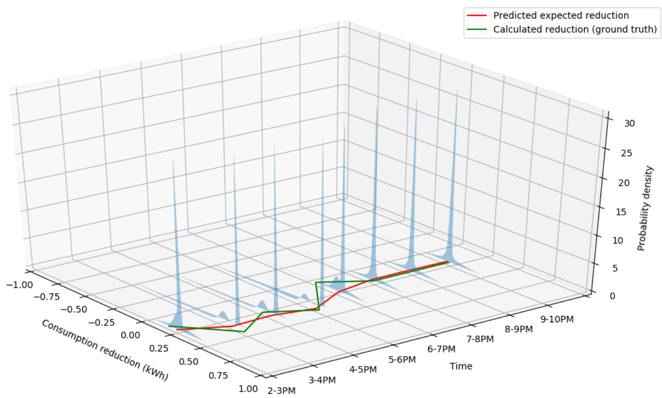


Fig. 5. Probability density of the predicted consumption reduction of Apt. 2 for a day (June 6, 2016) during summer.

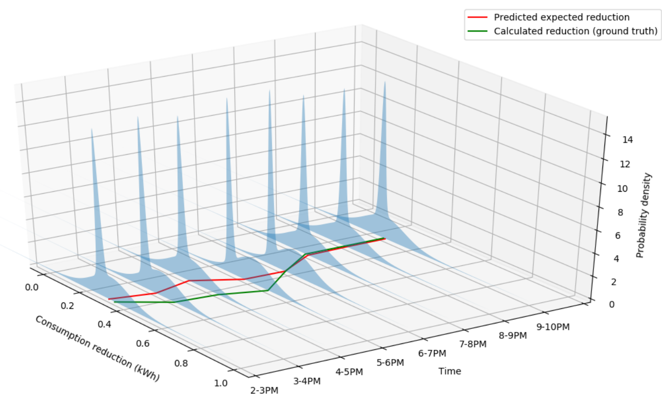


Fig. 6. Probability density of the predicted consumption reduction of Apt. 2 for a day (Jan 6, 2016) during winter.

results indicate that the RNN is able to learn the reduction capacity of a customer on an hourly and daily basis, while the MDN is able to generate probability distributions that give more insight into the reduction capability. A mixture of multiple Gaussian distributions provides a better estimate of the DR potential as opposed to a single distribution. Our proposed model depends on the customers' consumption sensitivity to the external temperature for obtaining an estimate of the load reduction in the event of a DR. To make the model more realistic, the training labels can be substituted by the actual values of the customer's load reduction. Due to the lack of availability of this data, we used the temperature sensitivity model to generate labels to show our model validity.

In order to maximize the profits and ensure the system reliability at all times, the DR aggregator must optimally choose the final set of customers to send the DR signals to. The uncertainty in customer behavior makes the optimal selection problem stochastic in nature. The results of the proposed model can contribute towards generating realistic scenarios for the stochastic optimization problem. In the future work, we will extend the application of this model to the stochastic customer selection process.

References

[1] D. P. Zhou, M. Balandat, and C. J. Tomlin, "Estimating treatment effects of a residential demand response program using non-experimental data," in *Proc. IEEE International Conference on Data Mining Workshops*, New Orleans, LA, USA, November 2017, pp. 95–102.

[2] U.S. Department of Energy, "Benefits of demand response in electricity markets and recommendations for achieving them," Tech. Rep., February 2006.

[3] S. Mohagheghi, F. Yang, and B. Falahati, "Impact of demand response on distribution system reliability," in *Proc. IEEE PES General Meeting*, San Diego, CA, USA, July 2011, pp. 1–7.

[4] P. C. Annika Todd and C. Goldman, "Residential customer enrollment in time-based rate and enabling technology programs," Lawrence Berkeley National Laboratory, Tech. Rep., June 2013.

[5] Public Utilities Commission of the State of California, "Resolution E-4728: Approval with Modifications to the Joint Utility Proposal for a Demand Response Auction Mechanism Pilot Pursuant to Ordering Paragraph 5 of Decision 14-12-024," Tech. Rep., July 2015.

[6] J. Kwac and R. Rajagopal, "Data-driven targeting of customers for demand response," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2199–2207, September 2016.

[7] F. Rahimi and A. Ipakchi, "Demand response as a market resource under the smart grid paradigm," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 82–88, June 2010.

[8] Q. B. Dam, S. Mohagheghi, and J. Stoupis, "Intelligent demand response scheme for customer side load management," in *Proc. IEEE Energy 2030 Conference*, Atlanta, GA, USA, November 2008, pp. 1–7.

[9] J. S. Vardakas, N. Zorba, and C. V. Verikoukis, "A survey on demand response programs in smart grids: Pricing methods and optimization algorithms," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 152–178, July 2014.

[10] Y. Chen, P. Xu, Y. Chu, W. Li, Y. Wu, L. Ni, Y. Bao, and K. Wang, "Short-term electrical load forecasting using the support vector regression (SVR) model to calculate the demand response baseline for office buildings," *Applied Energy*, vol. 195, pp. 659–670, 2017.

[11] D. Zhou, M. Balandat, and C. Tomlin, "Residential demand response targeting using machine learning with observational data," in *Proc. IEEE 55th Conference on Decision and Control*, Las Vegas, NV, USA, December 2016, pp. 6663–6668.

[12] J. Han and M. A. Piette, "Solutions for summer electric power shortages: Demand response and its applications in air conditioning and refrigerating systems," Lawrence Berkeley National Laboratory, Tech. Rep., January 2008.

[13] J. Kwac and R. Rajagopal, "Demand response targeting using big data analytics," in *Proc. IEEE International Conference on Big Data*, Silicon Valley, CA, USA, October 2013, pp. 683–690.

[14] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[15] C. M. Bishop, "Mixture density networks," Neural Computing research Group, Aston University, Tech. Rep., February 1994.

[16] K. O. Ellefsen, C. P. Martin, and J. Tørresen, "How do mixture density rnns predict the future?" *CoRR*, vol. abs/1901.07859, February 2019. [Online]. Available: http://arxiv.org/abs/1901.07859

[17] W. Wang, S. Xu, and B. Xu, "Gating recurrent mixture density networks for acoustic modeling in statistical parametric speech synthesis,," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Shanghai, China, March 2016, pp. 1–5.

[18] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht., "Smart*: An open data set and tools for enabling research in sustainable homes," in *Proc. ASM Workshop on Data Mining Applications in Sustainability*, Beijing, China, August 2012, pp. 1–6.