Regression and Bayesian Network in Electricity Demand Response

Jiqian Dong

***Abstract*— As sensing technology develops, more and more smart meters are installed for electricity data collection in the urban area. As a result, data driven energy management has become one of the key issues in smart city and other sustainable proposals. In this project, we developed a method for predicting the electricity consumption within a specific region of the city as well as investigating the correlation of the peak electricity demand between different regions. To be more specific, we integrated the data of civic geographic location, weather condition, social and population information to make the within-regional prediction using linear regression (LR) and Bayesian Ridge Regression(BR). After that, we built the Bayesian Network to further investigate the between-regional correlation in electricity demand peaks.**

I. INTRODUCTION

Since electricity can not be stored, once it is generated from the power plant, it is either consumed (transfer into the useful work by appliances) or wasted (transfer into heat and emits to the air). Therefore, maintaining the balance between supply and demand is an essential part for reducing waste and enhancing sustainability. Demand response (DR) is such a change in the power consumption of an electric utility customer to better match the demand for power with the supply. It's main task is to predict customer consumption of electricity dynamically and generate the equivalent amount of electricity accordingly.

In this project,we build data driven method for predicting the electricity demand with the help of modern analysis approaches such as python and machine learning algorithms such as Linear Regression and Bayesian Regression. Meanwhile, we can model the electricity consumption correlation between different regions since a great number of probabilistic graph-based models (e.g., Bayesian networks) have been developed recently to describe the behavior and interaction of a system.

Our main work can be summarized into 2 sections:

1. Electricity demand prediction within-region. We created both Linear Regression and Bayes Ridge Regression model to predict the hourly load consumption using following predictors including daily average temperature, daily temperature difference, hourly precipitation, daily air quality and hourly social pulse. Besides, we define the threshold for recognizing the demand peak and predict the probability of the demand that overpasses the threshold.

2. Electricity consumption correlation recognition. In this section, we mainly want to solve the question of what if scenario. For example, if the region A is going to have a "high" demand due to some events, what the probability of the adjacent region B's electricity consumption is also in "high" state. This is achieved by building Bayesian Network.

II. DATASETS AND DATA PREPROCESSING

A. Raw dataset

The datasets we collected to achieve these 2 goals are from Open Big Data collected by Telecom Italia. These datasets are specially designed to stimulate the creation and development of innovative technological ideas in the Big Data field. The raw data we use are listed as following:

- City Grids: In this dataset, the urban area has been meshed into grids. Each grid is a polygon that consists of multiple vertexes on the map and has an unique cell ID. The coordinates, presented in longitude and latitude are defined accordingly.
- Temperature: The temperature dataset contains the maximum and the minimum temperature of a single day collected by the weather station. The station has geographical coordinates that can be further mapped into the grid.
- Social Pulse: The social pulse is a dataset containing the people's activity on twitter. It is represented as the time stamp of people sending pushes.
- Precipitation and air quality: Both the precipitation and air quality dataset records the values within grids in one hour resolution. All the values are provided in categories, which are associated with different levels of precipitation volume or air pollution.
- Current: Current dataset provides information about the current flowing through the electrical grid of the Trentino province. It composes of Customer site dataset and Line measurement dataset. The Customer site dataset provides the number of customer site that a distribution line is connected within a grid. The Line measurement dataset provides the amount of current flowing through the lines in every 10 minutes intervals.

The visualization of the dataset is displayed in Fig 1.

B. Data Preprocessing

Since the sample resolution and spatial representation of the raw data in all above categories are different. Our major work in data preprocessing lies on synchronizing the time and mapping the geolocation where the data is collected to the cell ID.

For temperature data, we mainly calculated daily the average and difference from the maximum and minimum and mapped it to the grids using the coordinates of the weather station. For social pulse, since its active extent indicate the population density of a region, we first map the coordinates into the corresponding grid cell. Then calculate the total number "activities" by taking the summation hourly.

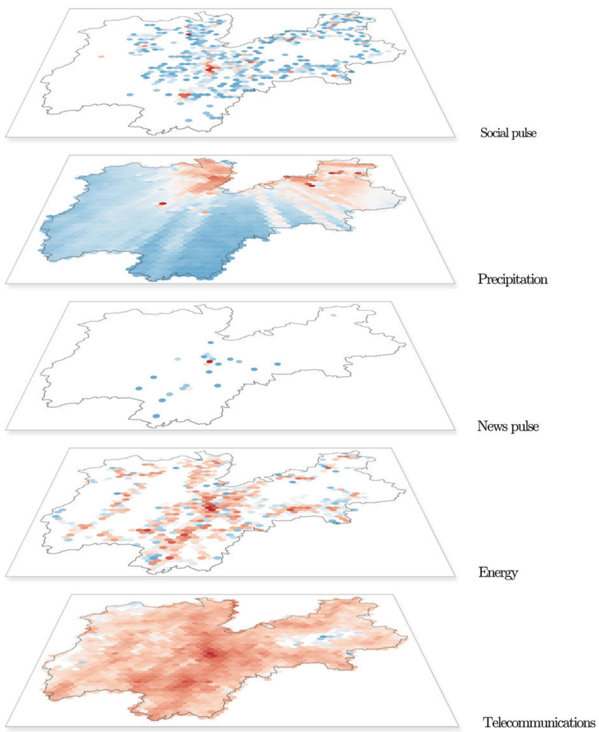

Fig. 1. Data Visualization

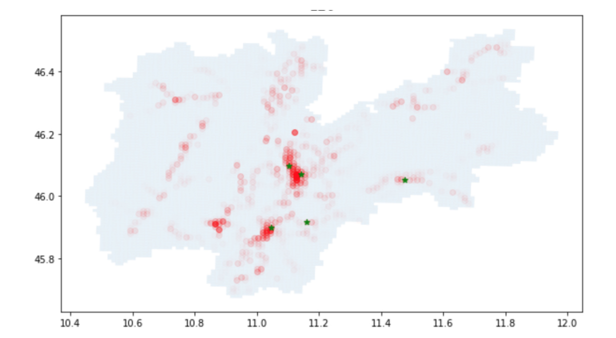

Fig. 2. Electricity consumption of all grids in 20:00. Green stars represent the chosen grids.

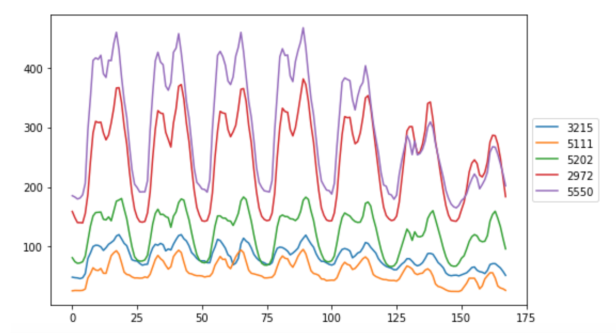

Fig. 3. Weekly electricity consumption of grids

For air quality and precipitation data, there are common that data might be missing for several hours because of sensor failures. Our solution is to make the missing data as another category, and do one-hot-encoding for all the categories.

For current data, since we are only interested in electricity consumption within grids, we transformed the line measurement data to grid measurement. From the dataset, we know that each line pass through multiple grids and contains different number of customer sites in each grid, and each grid might be associated with multiple lines. We make an assumption that each line distributes its power to the grids according to the number customer sites this grid is connected to this line. Based on that, we calculate the ratio a distribution line distribute its power and acquire the electricity consumption within grids.

Finally, after we synchronized the time of each dataset, we merged the temperature dataset, adjusted precipitation and air quality dataset into current dataset, using the grid number as merge index.

III. Method

As the entire dataset contains millions of cells, it is impossible to do the regression and network analysis upon all the cells. We selected 5 typical cells as our objective cells (shown in Fig 2) based on the absolute demand values (selected the largest ones) as well as the correlation between consumption pattern (shown in Fig 4.)and geographical distances.

Firstly, We used linear regression or Bayesian linear regression to predict the electricity consumption within a grid. In here, we get the probabilities that the electricity consumption is in normal state(temporal pulse less than a threshold) or in high peak(temporal pulse more than a threshold). Secondly, we used the electricity consumption state we predicted in the first phase, to infer the state of another grid, using Bayesian network.

A. Bayes Ridge Regression & Linear Regression

In this phase, we care about how the features within a grid influence the electricity consumption. Since we have 5 interested grids that are different from each other, we build models for each grid separately.

The weekly pattern of the load curve in chosen grids are shown in Figure 3. We decomposed the electricity consumption into a weekly pattern plus temporal as equation 1. The weekly pattern of a grid is decided by the intrinsic properties of that grid while the temporal pulse is associated with temporal features like temperature(T), social pulse(s) and precipitation(p).

$$u(t,T,s,p) = u_0(t) + u_{pulse}(T,s,p) \qquad (1)$$

We calculated the weekly pattern for each grid by averaging all the electricity consumption of that grid in that time. The reason is that we could marginalize all the parameters that cause the temporal fluctuations of the electricity consumption. Without additional information except time, we tend to assume the base pattern as current electricity consumption.

We used our linear regression model and Bayesian linear regression model to simulate the temporal pulse as equation 2. We constructed a limit state function 3 that compares the current electricity consumption with a threshold. If $g(u)$ is larger than 0, we consider the current state as normal, otherwise we consider the current state as in peak.

$$\begin{aligned} u_{pulse}(T,s,p) &= f_T(T)\theta_T + f_s(s)\theta_s + f_p(p)\theta_p + b + \varepsilon \\ &= \theta^T x + \varepsilon \end{aligned} \tag{2}$$

$$\begin{aligned} g(u) &= u(t,T,s,p) - threshold(t) \\ &= u_0(t) + u_{pulse} - threshold(t) \end{aligned} \tag{3}$$

1) Linear regression: In Linear regression model, we assume the error term follows a Gaussian distribution: $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. In the training phase, we calculated the coefficients of linear model using normal equation: $\theta = (X^T X)^{-1} X^T Y$. In the prediction phase, the model predicts the temporal pulse of electricity consumption using the inputs and coefficients θ. Finally, we compares the u with threshold in the limit state function. Since the output of linear regression does not hold probability meaning, we assume the probability of peak state as 1 if g is larger than 0.

2) Bayesian linear regression: In Bayesian linear regression model, we assumed the Gaussian prior for error term ε ($\varepsilon \sim \mathcal{N}(0, \sigma^2)$) and coefficients θ ($\theta \sim \mathcal{N}(0, \sigma_\theta^2 I_d)$). Based on that, our prediction of y is as equation 4. For now, we assume $a = \frac{1}{\sigma^2}$ and $b = \frac{1}{\sigma_\theta^2}$.

$$Y_i \sim \mathcal{N}(\theta^T X_i, \frac{1}{a}) \tag{4}$$

In the training phase, the distribution of θ updates with the data to maximize the probability. After calculation, the posterior of θ is given as 5

$$P(\theta|D) \sim \mathcal{N}(\mu, \Lambda^{-1}) \tag{5}$$

$$\Lambda = aA^T A + bI_d$$

$$\mu = a\Lambda^{-1} A^T Y = (A^T A + \frac{b}{a} I_d)^{-1} A^T Y$$

$$P(y|x,\theta) \sim \mathcal{N}(\mu_{theta} x, a + x^T Lambda_{-1} x). \tag{6}$$

In the prediction phase, given the x, we could get the distribution of the electricity consumption, which should also follow a Gaussian distribution 6

After we acquire the distribution of y, we could calculate the distribution of the limit state function, and thus get the probability that the current electricity consumption is in normal state and peak state.

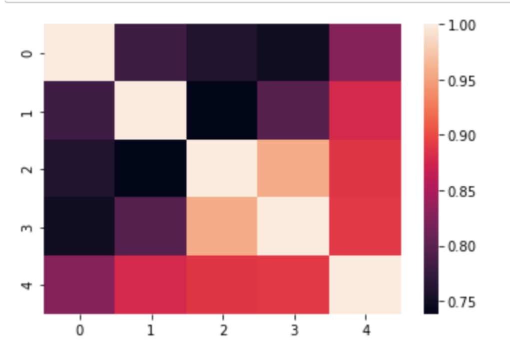

Fig. 4. The correlation between objective grids

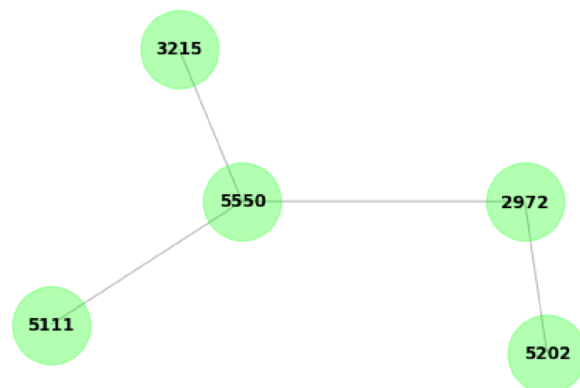

Fig. 5. Network graph for objective grids

B. Bayesian Network

As mentioned before, our network is connected base on the correlation matrix between the grids. However, this network is not a direct graph as there are no arrows in between. According to the definition of Bayesian network, which is a probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG).Therefore we need to further define the direction of each path. In this project, the Bayesian Network is defined dynamically by given 2 nodes, denoting the start node and the end node. The start node will be the one that the probability of electricity demand in “high” or “low” stage is given, and the corresponding output will be the probability of the same issue at the end node.

For each Bayesian Network, there are 3 main components that are indispensable for the inference, the graphical topology, the conditional probability table (CPT) and the initial (marginal) probability. For our network,

1) Graphical topology: The graphical topology of the network is defined dynamically given the start node and end node. Then the program will automatically select the shortest and acyclic path to link them together. For example, if the given start node and end node are “5111” and “5202” respectively, the network will be the direct path link “5111”,“5550”,“2972” and “5202” sequentially.

2) Conditional probability table: The conditional probability is defined for each path, for example, the CPT for path "2972" to "5550" is shown in table I. The first column represents the given condition of the start grid, the entry in each row depicts the probability of each state in the end grid given the start grid. Thus each row sums to 1 according to the uniformity property of probability.

TABLE I
AN EXAMPLE OF A CPT FOR PATH '5550' TO '2972'

Condition	'5550' high	'5550' low
'2972' high	70%	30%
'2972' low	15%	85%

3) The initial marginal probability: The initial marginal probability is the output from the Bayesian Regression model, as mentioned in section 1. It can be interpreted as the probability of start node in high or low electricity consumption.

4) Bayesian Network learning: The learning of Bayesian Network is based on the Bayes formula and maximum likelihood estimation. Consider a problem of 2 nodes,B is the start node and A is the end node. We need to calculate the The combination of these 2 method yield to the "counting method", which is shown in the following equation.We only need to count the frequencies of all the observations and take the fraction as shown in equation 7.

$$\begin{aligned}\Pr(A = high|B = high) &= \frac{\Pr(A = high \wedge B = high)}{\Pr(B = high)} \\ &= \frac{\sum_{obs} \delta_{A=high \wedge B=high}(obs)}{\sum_{obs} \delta_{B=high}(obs)} \qquad (7)\end{aligned}$$

where δ is the indicator function see 8:

$$\delta_A(x) := \begin{cases} 1 & \text{if } x = A, \\ 0 & \text{if } x \neq A. \end{cases} \qquad (8)$$

The learning process of Bayesian Network is done by feeding the deterministic data from the first section.

5) Bayesian Network inference: Since a Bayesian network is a complete model for the variables and their relationships, it can be used to answer probabilistic queries about all the nodes. In this case, the network we consider is the simplest case: from one node to another node. Thus the inference work can be done by adopting the chain rule and total probability formula (shown in equation 9). Now, with any input at any node by giving the probability of its consumption in "high" or "low" state, we are able to infer the corresponding probability of the objective node.

$$\begin{aligned}\Pr(A = high) &= \sum_{B \in (high,low)} \Pr(A = high, B) \qquad (9) \\ &= \Pr(A = high \wedge B = high) \\ &+ \Pr(A = high \wedge B = low) \\ &= \Pr(B = high)Pr(A = high|B = high) \\ &+ \Pr(B = low)Pr(A = high|B = high)\end{aligned}$$

IV. RESULT

We split the dataset into training and testing set randomly with testing scale as 0.2. We used training set to train our regression and Bayesian network. Then, we used the testing set to evaluate the performance of our models. Currently, we only used daily temperature mean and temperature difference, and hourly precipitation as our features when we did regression. Further, we would consider social pulse.

We used the Root Mean Square to evaluate the regression effects of our regression models. For Bayesian linear regression, we used the mean of the posterior distribution as the results. The Table II illustrates the RMS score for two models. It could noticed that the two models performs similar, and the Bayesian linear regression model(BL) is slightly better than the linear regression(LR) model. And this advantage is more obvious when the training sample is small, since BR models assumes a prior for parameters and update this prior during training. However, both the two models does not provide a good regression effects. It might because we did not use the social pulse data that actually explains the temporal fluctuation best.

TABLE II
COMPARISON OF RMS SCORES FOR TWO MODELS

Grid number	BR	LR
'3215'	7.70	7.72
'5111'	8.41	8.42
'5202'	7.36	7.37
'2972'	12.64	12.71
5550	23.04	23.13

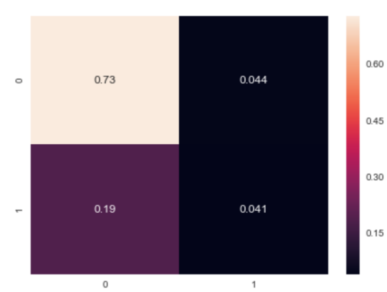

Fig. 6. Confusion matrix of Bayesian Network predicting

We used accuracy to evaluate the performance of our final model. To be specific, we applied the LR & BR model to acquire the state probability for one grid. And then we used this probability to infer the state of another grid. The final state will be either "normal" or "high" depending on which probability is larger. The Figure 6 illustrates confusion matrix of the predicted results when we use BR as regression model. The diagonal of the matrix define the correct classified label. The correct prediction of "normal" has the probability of 0.75 and the while it only has the probability of 0.041 to correctly recognize the "high" stage. Several reasons might account for this. Firstly, it might because we predict the state of testing node based on the assumption that we only know the features of one node. The information of multiple grids or the features of the testing node may significantly improve our prediction. Secondly, it might because of the inaccurate in the Linear regression models, the error will propagate from one model to another and the total loss is amplified. Thirdly it might because we didn't not utilize the transition relationships that associate states in neighboring time. Finally, the imbalanced data with much more entries for the "normal" state than "high" state might make the final prediction unbalanced.

V. CONCLUSIONS AND FUTURE WORK

As illustrated before in the regression work, BR and LR model have similar result and performance in predicting the absolute value of electricity demand within grid, and the overall accuracy is acceptable. However, from the result of Bayesian Network,it is obvious that the model has an unbalanced better accuracy in predicting "low" state over the "high" state. The overall performance is unsatisfactory.

As for the future work, we can do several modifications to enhance the accuracy of this combined model. First, for the regression model, we can modify the current feature space or create more realistic and informative features to enhance the accuracy. This can be done by doing more data cleaning work such as temperature interpolation or adding time transition features. For the Bayesian Network Model, since the current network has only "one way" topology (from one single node to another), we can improve the accuracy by adding "double ways". That is only simply use one node to predict another, we can improve the CPTs by adding multiple predictors. For example, we can consider the conditional probability node "5550" is "high" given the joint probability of both "2972" and "3215" are "high". Furthermore, we can by feeding the network with more generalized data to prevent it from biased.

REFERENCES

[1] Zhang, N. L., & Poole, D. (1996). Exploiting causal independence in Bayesian network inference. Journal of Artificial Intelligence Research, 5, 301-328.

[2] Barber, D. (2012). Bayesian reasoning and machine learning. Cambridge University Press.