Building Energy Consumption Forecast

Author: Taiyu Chen
Adviser: Prof. Andrew R. Cohen

Abstract

Predicting energy behavior of a building is great practice for unknown future forecasting. There are various features related to the office building energy consumption. But it is still a challenge to design an effective prediction learning machine model that gives us a great performance of forecasting the current hour building energy usage. In this study, we will use RBF-SVMs to classify features in order to make accurate assumption of energy behavior of Nesbitt Hall at Drexel University. We first create a RBF-SVMs model using previously established sets of features. At the end of discussion, we will explore more about feature selection to select suitable features and integrate them to calculate the energy forecast for current hour from hundreds of features. In the result, we could see the test of this RBF-SVMs model comparing to real-time energy behavior, and the performance curve of accuracy of this model for 150 days.

Introduction

Global warming is a hot topic in this recent years, since the environment has become more and more unfriendly to human beings. The carbon pullon is the main reason for it. Every Year a lot of the energy that the buildings consumes is wasted through heat loss and inefficient technology, which can lead to increased carbon pollution. An accurate building energy consumption forecast tool could play an important role in building design and real-time control which have a great impact on energy efficiency. We are going to conduct a research of building energy behavior forecasting in Nesbitt hall which is an old building at Drexel University. It contains one chiller and two heating exchange system which are the most electric-consuming equipments. There are about 540 sensors have been deployed in Nesbitt Hall. Data are collected every five minutes from these sensors. Each data from a sensor represent a factor which is a candidate to be selected to join a subset of features to predict the building energy consumption. Furthermore, except for these 540 sensor, there is a special sensor is used to record the overall electric power consumption every hour. The data from this sensor can be used as ground truth to evaluate the performance of each subset of features. In this research, we are going to build a acceptable method to predict the current hour energy usage based on the various features we collected from previous hour. In this study, we will try to use a machine learning technique called Support Vector Machine(SVM) to build a successful prediction model.

Methods

There are a lot of computational method that can be used to build this model. Generally, they can be divided into two categories: unsupervised and supervised methods. In this study, we focus on supervised methods. The supervised learning means that training data is a series of labeled pattern, where each pattern is a collection of features that is labeled with the correct output corresponding to that feature set [1]. We could understand it in this way that the algorithm takes given features as inputs and outputs the particular label, and must apply what it "learns" from the training dataset to predict the outputs (labels) for another testing dataset. Supervised learning can also be further broken down into classification and regression problems.[1] In classification problems, there are a set of outputs (patterns) that a feature set can be labeled as, whereas the output can take on continuous values in regression problems.[1] In this study we consider the problem of the hourly energy usage forecasting as a classification problem. SVM is the best binary classifiers we could find so far. SVMs are based on the principle of structural risk minimization which aims at restricting the generalization error at the lowest level.[2] It creates a decision line or hyperplane such that most points in one category fall on one side of the line or hyperplane while most points in the other category fall on the other side of the line or hyperplane. Consider an n-dimensional feature vector x = (X1,...,Xn) [4]. We could write down the liner line or hyperplane:

$$\beta 0 + \beta 1 X1 + ... + \beta n Xn = \beta 0 + \sum_{i=1}^{n} \beta i Xi = 0;$$

Then elements in this specific category will be such that the sum is greater than 0, while elements are not in this specific category will have the sum be less than 0.

Let's make a new function:

$$y(x) = \beta 0 + \sum_{i=1}^{n} \beta i Xi, \ \ y \in \{-1, 1\} \ ;$$

So *y(x)* could be the determining function to see if a specific feature can be can be classified into a specific pattern.

We also can rewrite this *y(x)* using inner products, because that will be earlier for Matlab to computer:

$$y = \beta 0 + \sum \alpha i y i x(i) * x$$

Note that we have sussfucally maken the inner product by its label.

Through this optimal hyperplane, we maximize the distance from the plane to any point, which is known as the margin. The maximum margin hyperplane can best splits the data.If we still are not satisfied with the performance of this model, we can add error variables ε1...εn and keep their sum below a certain value. The crucial meaning is that only the points very closest to the boundary have the impact for the hyperplane selection; all others cannot be selected.. These points are support vectors, and the hyperplane is the Support Vector

Classifier (SVC) in this case. In this study of forecasting of energy, we don't put extra weights $\varepsilon 1 ... \varepsilon n$ to keep their sum more expected because it is already enough for us. SVCs are limited since they are only linear boundaries.[5] We could use RBF-SVMs to fix it because kernel functions are able to map the inputs into a higher-dimensional space and linearly classify inputs in that space. The RBF was selected as a kernel training function due to the high regression precision. [10] The RBF kernel $K(x(i), x) = \exp(-\gamma||x(i) - x||^2)$ is one of the most popular kernel functions and great option for us in this study.

We could rewrite the hyperplane function again by plug in with the RBF kernel K(x, x0):

$$y(x) = \beta 0 + \sum \alpha i y i K(x(i), x) \quad y \in \{-1, 1\}$$

In this study we will use RBF-SVMs first since we have already successfully build the mathematics model of RBF-SVMs for it.
Based on the dataset from those 540 sensors. We have already had *d* features (inputs) for 24 hours in a day of energy usage in a matrix D1. Since we already collect more than one year dataset. We can use a few days for training and the rest of days for testing. But in this case, we only choose two days for this experiment. So we could combine Day1 and Day2 matrix into a new matrix D2 that includes 48 hours energy usage record. Let's say D2 is the matrix D(all). We use the first 1..24 hours for training and 25...48 for testing of our algorithms. We use x as a variable to define the xth hour in that day. We use *y(x)* to denote the *d* dimensional feature vector for a certain pattern. For each hour, we define *h(x)* as the value of energy usage hourly from the sensor, and define the target output based on whether the hourly energy usage *f(t) as the change* that increased or decreased that drop in which intervals compared to the previous hour:
we also classify the increment and decrement into six intervals (patterns):

$$\{+2, +5, +8, -2, -5, -8\}$$

In the trainning set, we define that

f(t)=2 if 0<h(x)-h(x-1)<=2
f(t)=5 if 2<h(x)-h(x-1)<=5
f(t)=8 if 5<h(x)-h(x-1)
f(t)=-2 if -2<=h(x)-h(x-1)<0
f(t)=-5 if -5<=h(x)-h(x-1)<-2
f(t)=-8 if h(x)-h(x-1)<-5

We define the trainning set using Z= $\{X1, f\}$. In this case, D1 are rows from the first 24 hours of the matrix X and f is the prediction target output for the first 24 hours. We use this training set not only to select the best suitable subset of features but also to training our classifiers.

However, there is a problem coming: SVM is internally a binary classifier. This means that it can only separate two labels. But we have at least six patterns need to classifier so far. In order to fix this problem, we have to use SVM algorithm for six times. Each SVM represent a case of pattern. We basically create *ensemble* of SVMs, which vote for particular patterns. In this case, we have to create 6 models. The i'th SVM is trained to recognize if the object can be labeled the ith pattern or not. If it can, then the y(x) of i'th SVM is 1, otherwise is -1 because $y \in \{-1, 1\}$; Then, when it comes to classification, the program asks each SVM if it recognizes the pattern.

Before we use RBF-SVMs, we also have to manually select the best subset of features from these 540 features. We chose the most six factors data from these 540 data sets to become the most determinant factors. But they are not features yet!

Table 1. Factors selected by domain knowledge.

Outdoor air temperature
Outdoor air relative humidity
Chilled water flow rate
Control chiller status
Between-coil temperature
Supply air temperature setpoint

We need to find the change of the data from these six factors above in order to build feature space.

Since these six factors are from those 540 sensors, the data of these six factors must be collected every five minutes. Thus, the sensor can generate 60/5=12 data for each hour. We use the current hour mean value $\mu 1$ of 12 data minus the previous hour mean value $\mu 2$ for each factor as the feature x from this factor. In this case, we will create a six-dimensional feature vector for each hour:

$$\mathbf{X=[x1,...,x6]}$$

And we put the first day hourly feature vector into the training feature matrix X1 to train the six SVMs classifiers. And we use the second day feature matrix X2 as testing dataset to test these six SVMs classifiers in order to see if they are able to give us the approximate target output $\{2, 5, 8, -2, -5, -8\}$;

Besides, in order to evaluate the performance and accuracy of the training machine, we test it by the following formulation:

$$acc = \frac{1}{48-24} \sum_{x=25}^{48} ((h(x) - h(x-1)) - f(x))\,\hat{}\,2$$

The accuracy is the 1-(sqrt(acc) /($\frac{1}{48-24} \sum_{x=25}^{48} ((h(x) - h(x-1))$)

Experimental Result for RBF-SVMs method

Table 2. Test dataset result

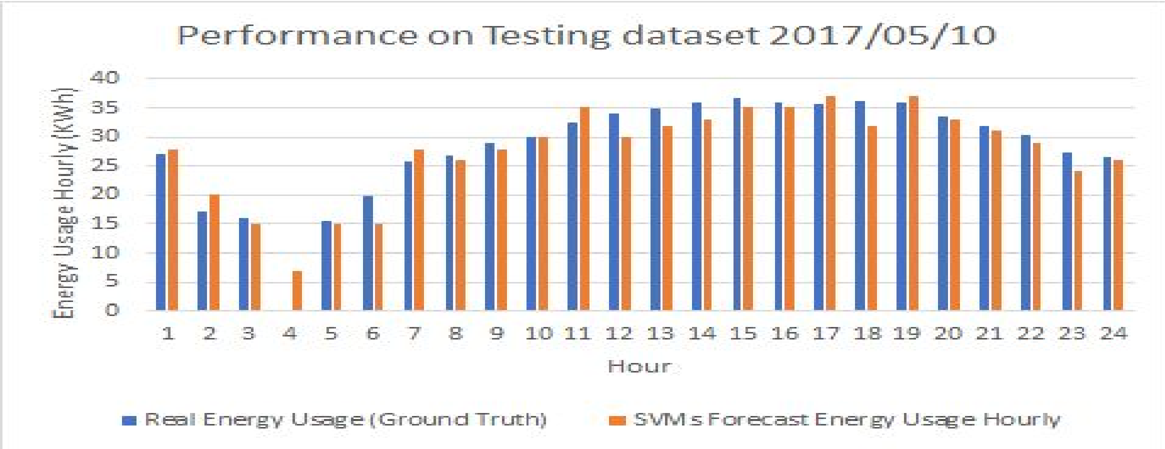
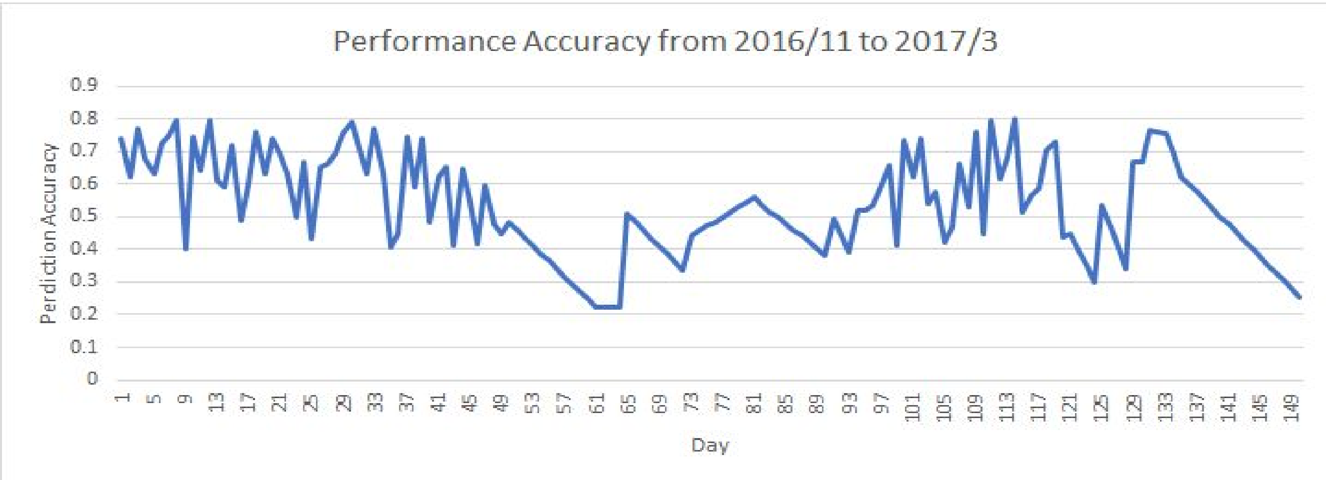


Table 3. the prediction accuracy for 150 days test dataset



DataSets

We use 6 features from the dataset obtained by 540 sensors owned by Drexel University Engineering Group [3]

Related work

There are some previous work on this topic but they are more interested in traditional statistics ways and used more than 15 features in their assessment of traditional regression analysis, decision tree and neural networks. These works can be found in [7–9]. They all didn't try to build a learning machine or model for forecasting in the future use.

Discussion and Conclusions

The basic idea of office building energy prediction is to extract mathematical models from historical performances dataset and to apply these models to predict the energy usage behaviors in the unknown future. It can become a powerful tool for achectrues designing building and for maintenance team for troubleshooting. Comparing to other predictive approaches such as engineering calculations, measuring and monitoring, and physical simulation that we have learned from foundational classes, the advantages of supervised learning methods are obvious. Since the prediction is based on solid and realistic historical dataset, the performance of prediction is able to be accurate. Furthermore, it is comparatively easy to apply the models for prediction once the models have been built and proven work well. We even can make this mathematical models into program, function and package in computer. When people need to use them, they just simply call them with appropriate inputs in coding. Even for those people without any data science and machine learning skills and training could still run and use this models. Many people could benefit from the convenience of this models.

From the result of the experiment for RBF-SVMs method, we could see that the performance of prediction for 150 days are not very stable. The accuracy of the prediction could be not only acceptable for some days, but also unexpected low for some other days during these 150 days test. We assume that the reason why it happens is that we didn't give the RBF-SVMs model enough training and surprising opportunity. We only use one day for training the RBF-SVMs classifiers which is the dataset on 2017/05/10. But in real situation when the season change, the building is very likely to have distinct energy behavior from other seasons. Also, it is also possible that some features are only active in some days but silent for the rest of days of a year. In the future, if we redo this project again, we should consider use more days as training dataset. But it will also cause a new challenge to set the suitable weights for each β and the bandwith $-\gamma$ of the kernel function. It is very time consuming to find out a appropriate value for the weights of each feature vector when we design the RBF-SVMs model.

Additionally, we only use six features to create the featurespace for training by domain knowledge. We don't even know whether six features are enough or not. Feature selection often has the biggest impact on a machine learning model's accuracy. We will try other feature combination or add more to the feature list. Future work would involve adding features related to weather related variables and related to Cooling coils. We have also learned feature selection from class. We could use SVM-RFE to output a ranked list of the features from those 540 potential features. And then we use the Pearson correlation coefficients to determine the relationship between these features and rearrange the ranked list of the features in order to remove the those features that have similar performance to other features.[10] Finally, we could build the model with the best suitable subset features. In conclusion, real situation is always more complicated than experiment with ideal features. We may only predict the energy behavior in a centain range with the same model.

Reference

[1] J. Brownlee. A tour of machine learning algorithms

[2] V.N. Vapnik. The nature of statistical learning theory. Springer-Verlag New York, Inc., New York, NY, USA, 1995.

[3] Drexel University Engineering Group, “point_indices_Nesbitt,” Between 2016/11 to 2017/05

[4] M. Halls-Moore, “Support vector machines: A guide for beginners,” September 2014. [Online].Available:http://www.quantstart.com/articles/Support-Vector-Machines-A-Guide-for-Beginners

[5] M. Espinoza, J. Suykens, and B. Moor. Fixed-size least squares support vector machines: a large scale application in electrical load forecasting. Computational Management Science, 3(2):113–129, April 2006.

[6] ScienceDirect, “Overview of Support Vector Machine in Modeling Machining Performances, ” 28 December 2011

[Online].Available: https://www.sciencedirect.com/science/article/pii/S1877705811054993

[7] B. Dong, C. Cao, and S.E. Lee. Applying support vector machines to predict building energy consumption in tropical region. Energy and Buildings, 37(5):545–553, 2005.

[8] A.E. Ben-Nakhi and M.A. Mahmoud. Cooling load prediction for buildings using general regression neural networks. Energy Conversion and Management, 45(13-14):2127–2141, 2004.

[9] S.L. Wong, Kevin K.W. Wan, and Tony N.T. Lam. Artificial neural networks for energy analysis of office buildings with daylighting. Applied Energy, 87(2):551–557, 2010.

[10] ScienceDirect, “Stable feature selection for clinical prediction: Exploiting ICD tree structure using Tree-Lasso,” December 2014.

[Online].Available: http://www.sciencedirect.com/science/article/pii/S1532046414002639