FORECASTING CUSTOMER'S ENERGY DEMAND USING MACHINE LEARNING

SAIFUL ABU

Department of Computer Science

APPROVED:

_____

Christopher Kiekintveld, Chair, Ph.D.

_____

M. Shahriar Hossain, Ph.D.

_____

Paras Mandal, Ph.D.

_____

Charles Ambler, Ph.D.
Dean of the Graduate School

to my

MOTHER and FATHER

with love

Abstract

Accurate electricity demand forecasting is an important problem as the failure to do so may be costly for both economic and environmental reasons. Power TAC simulation system provides a no risk platform to do research on smart grid based energy generation and distribution. Brokers are important components of the system. The brokers work as self-interested entities that try to maximize profits by trading electricity in various markets. To be successful, a broker has to forecast the electricity demand about its customers as accurately as possible, otherwise it will operate ineffectively. This proposed forecasting method uses a combination of cluster and classifiers. At first, the customers are clustered based on their weekly average usage. After that, energy usage history and related weather related information are combined together to train classifier for the cluster. To forecast for a new customer, the proposed method needs at least a week's energy usage history of the customer. The system assigns the new customer to one of the clusters based on its electricity usage history. The classifier for that cluster will be used to forecast the customer. This approach produced 13 % error compared to 31% relative absolute error observed against the moving average baseline predictor. The Power TAC system has six different types of customer such as customers with demand shifting capabilities, customer with no demand shifting capabilities, electric vehicles, thermal storage, wind production and solar production. Previous approaches to demand forecasting treated all types of customers equally. This work shows that a good forecasting system should treat customers of different type differently, otherwise the system will experience more error.

Table of Contents

List of Figures

Chapter 1

Smart Grid and PowerTAC Competition

In this chapter, I will describe the smart grid [5] and Power TAC [14].

1.1 Traditional Electricity Distribution and Consumption System

In traditional power grids, there are three subsystems electricity generation, transmission, and distribution [5]. In electricity generation subsystem, the generator rotates a turbine in a magnetic field which generates electricity. The turbine rotates through the power of kinetic energy of water falling from a waterfall or a river with strong current, or from the energy of nuclear power plant, or energy received from burning coal or oil. Traditional energy generation system then transmits the electricity through transmission grid and electricity gets distributed through the distribution grid. This generation system is one way meaning a the electricity flow occurs from source node to consumption node only.

1.2 Smart Grid

In contrast to the traditional electricity generation system, Smart Grid are two-way [5]. So, any node in the distribution grid can produce electricity and push it to the distribution grid if necessary. The NIST report [5] states that the SG would make the electricity generation and supply robust against generator or distribution node failure, use renewable energy widely and efficiently, reduce greenhouse gas emission, reduce oil consumption by encouraging usage of electric vehicles, it will give customers more freedom to choose among energy sources. Smart grids will encourage usage of the electric vehicle as these vehicles have the ability to store power in a battery and transmit the power to the distribution grid if there is a necessity. The major challenge with the usage of renewable energy is it is uncertain. This uncertainty causes the ability to predict how much energy the SG can produce in a future time slot hard. The success of SG will need efficient methods to predict energy production [22].

1.3 Smart Grid and Renewable Energy

One of the major focus of Smart Grid will be using renewable energy. There are challenges involved with using this abundant source of energy [25]. People are already showing strong motivation to use renewable energy as indicated by the statistics that 20% of total energy is from the renewable sources which are second after coal 24%. Consumers are using renewable energy due to economic reward and environmental concern. A major challenge with renewable energy is the amount of the energy produced is greatly variable. Since the energy produced is volatile there must be a storage mechanism that balances out the surplus energy. The usage of rechargeable electric vehicles might serve the purpose of storage. Accurate prediction of the renewable energy might enable the electric car users to absorb surplus energy and push it back to the grid in peak hours if necessary.

1.4 Power TAC System

Power Trading Agent Competition (Power TAC) [14], [15], is a low-risk system that simulates a smart grid based energy system. This simulation system models a competitive and liberal energy trading market. The power TAC simulation has several components such as wholesale market, brokers, customers, distribution utility and weather service [15]. The brokers publishes tariff plans for electricity consumers and producers. It then buys electricity from the wholesale and balancing market to meet customers need. The system is trained on customers behavior of past years and uses real weather data. The following sections give a brief explanation of each component of the Power TAC. Figure 1.1 shows a block diagram of the components of the powerTAC simulation environment.

1.4.1 Broker

In Power TAC system, participants implement their broker logics. Here are the list of actions that a broker can take in Power TAC simulation -

- At any hour of the simulation, a broker can publish a new tariff. Each tariff is targetted to a specific category of the customers. The tariffs contain information about which category of customers it is targetted, expiration date of the tariff, signing bonus, penalty for early withdraw and rate of periodic payment.

- At any moment of time, a broker can modify its published tariffs. It can adjust payment rates and withdraw penalty. The broker can also revoke a tariff that is not profitable.

- To meet customer's demand, a broker takes part in the electricity auction in the wholesale market. There, it specifies how much energy it needs and how much it is ready to pay for. Based on the asks and bids from other brokers in the simulation, the broker's bid may or may not get cleared.

At any moment of time the brokers are aware of the following information -

- Participating brokers in the simulation

- Customers present in the simulation.

- Bootstrap data of customers and wholesale market.

- Published tariff in the simulation.

- Information about tariff modification or revocation

- Wholesale market clearing prices of last time slot.

- Energy transaction information of subscribed customers.

- Transaction of balancing market.

- Current bank balance of itself.

1.4.2 Customers

A customer represents an entity that buys energy from the brokers. A customer has the following attributes -

- An unique name.

- Number of individuals it represents. This number can range from one to several thousands.

- A power type that specifies which category of customers (producer or consumer) does it fall into. Producer and consumer categories have several subcategories.

A customer can take the following actions during a simulation -

- Evaluate available tariffs in tariff market.

- Subscribe or abandon a tariff. Customers try to maximize their economical gain so if there is a lucrative tariff in the market, a customer might try to subscribe to it.

- Generate meter reading based on produced or consumed energy. The system then sends this meter reading to the broker it subscribed to.

- Customers with demand shifting capabilities can shift their demand to favorable time slot.

1.4.3 Weather Service

The weather service broadcasts weather forecast of future hours and weather report of current hour to the brokers. The weather report contains information such as wind speed, cloud cover, temperature, day of week and month of week. Power TAC uses real weather data from the past that makes the simulation more realistic. Brokers can use these information to forecast demand for the weather sensative consumers. The weather information also makes it possible to forecast about renewable energy producers.

1.4.4   Markets and Distribution Utility

There are three different types of markets in Power TAC simulation, namely, wholsale market, tariff market and balancing market. The wholesale market is the bidding place for buying energy. Bulk energy producers and brokers take part in the wholesale market auction. Brokers can submit their bids for 24 future timeslots in the wholesale market by specifying the price it is prepared to pay. If the bid was successful, the broker receives its desired amount by paying the money. At each time slot, the system notifies the broker about the wholesale market clearing prices. Brokers publish their tariff plans in the tariff market. A tariff holds information about the pricing of the energy. Customers, upon analyzing available tariffs, subscribe to their mostly suited tariff plan. Balancing market represents the market from where the broker can buy energy in case of emergency. For example, if a broker has bought less amount of energy for a given timeslot and it finds it needs more energy then it can buy the necessary amount of energy from the balancing market. Usually, the balancing market transactions are costly for brokers than the wholesale market.

The distribution utility has two main objective. First, it supplies energy to the consumers from whole sale market and from the renewable producers. Secondly, it works as a default broker that publishes default tariffs at the start of the game. This makes sure no customer is ever out of energy. Other brokers are supposed to publish lucrative tariffs to attract customers.
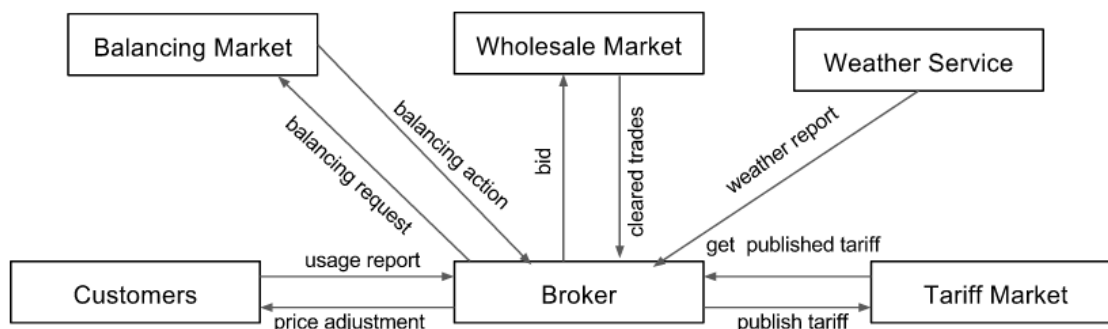
Figure 1.1: PowerTAC simulation environment.

1.5   Importance of Accurate Demand Forecasting in Power TAC

A broker has to make bids and asks in the wholesale market. The amount of electricity it asks depends on the demand forecast of its subscribed customers. If the broker fails to make accurate demand forecast, it will not be able to ask for proper amount of electricity. So it will end up asking more or less energy than the required amount in the wholesale

market. In this case, the broker will have to buy energy from the balancing market in a higher price or has to sell surplus energy in a lower rate. As a result, it will face monetary losses. This unwanted scenario can be avoided through demand forecasting as accurate as possible. This thesis investigates ways to find a better demand forecasting mechanism.

Chapter 2

Related Works

In this chapter, I have described different methods of energy load forecasting for long term and short term in the literature. It is hard to know the state of the art electricity consumption mechanism in Power TAC as most of the researchers did not publish their demand forecasting mechanis [16], [20], [28], [19]. So I mostly describe the works done for real world electricity demand forecasting mechanisms.

2.1   Variables in Electricity Demand

Studies such as [10], [7] and [4] have found that temperature has effect on electricity demand. The study in [7] was done in a region of Australia. It was found that, in a lower temperature the customers tend to use heaters and in a higher temperature they tend to use coolers. As a result, the increase or decrease of temperature from a certain point will cause the consumption of electricity to increase. In study [4], two demand forecasting models were proposed. One was univariate Auto Regressive Integrated Moving Average (ARIMA) and another one was univariate ARIMA model along with temperature depended transfer function. The model with temperature variable did better forecasting than the one without the temperature variable. On the other hand, the study in [3] showed that inclusion of temperature variable in forecasting model actually introduced more error in demand prediction. The aim of the study was to make forecasting about electricity usage of January based on past five years training data using a Support Vector Machine (SVM) forecaster. The reason behind of getting more error after including temperature variable may be because during Januray the temperature did not change much and the inclusion might have caused overfitting.

Weather variables such as wind speed and cloud cover has effect on electricity demand [10], [26]. As cloud cover increases, the demand for light increases too. The increased lighting demand causes increased electricity demand. The period where cloud cover was low, the electricity demand was also low [10]. High speed wind across wet walls help cool houses. High speed wind thus may cause reduced electricity demand due to reduced demand of air cooling [26].

In the survey article [6], the authors reported that the day of the week and the month of the year is highly correlated with customer's energy demand. The electricity load demand can be higher or lower based on the day of week. The weekends usually have different load demand pattern than the week days. Also, based on the hour of a given day, the load demand can be higher or lower too. The season also showed impact on electricity demand.

2.2 Electricity Load Forecasting Using Statistical Method

To make electricity load forecast, researchers have used statistical methods such as statistical average, Auto Regressive Integrated Moving Average (ARIMA) and exponential smoothing. Agent TACTEX'13 [29], the winner of the PowerTAC competition in 2013, used the statistical average to make electricity demand forecasting for an hour of a day of a week. In a week a customer has 24 * 7 = 168 hours or slots where it can consume electricity. TACTEX'13 kept track of average usage of 168 weekly slots for each customer. To predict a future time slot, their agent would look at which weekly slot the future time slot would fall in. Then the agent used that weekly slot's average usage as the forecast of the future time slot. [4] have used an ARIMA model for load forecasting. The ARIMA model uses both moving average and auto regression to forecast the demand. To make a forecast about a future time slot, the auto regression model uses some previously observed time slots values based on its degree. Moving average scheme would use the average of all the known time series data points to make a prediction about a future time slot. In the study [2], a short term load demand was proposed that uses several ARIMA models. For the combination of week day and temperature level, 16 short term load forecasting models were used. This scheme made better forecasting than a single ARIMA model. The authors in [13] used modified Halt Winter Exponential Smoothing for demand forecasting. The modified exponential method was cabaple of dealing with weekly and daily seasonality pattern present in the data.

2.3 Load Forecasting using Machine Learning

Support Vector Machine(SVM) proved to be an effective tool for load forecasting [27], [3]. In [3], the authors used SVM to forecast electricity demand of January based on past 5 years electricity consumption data. Separate SVM models for separate were proposed. SVM model trained with data from January was able to make better load forecasting. Artificial Neural Network (ANN) is another favorite load forecasting mechanism among the researchers [12], [23], [11]. The author [23], used ANN to model Prediction Interval (PI) to forecast renewable energy forecasting. The author [12] used ANN to figure out the time horizon suitable for solar energy production. [21] the authors used various machine learning techniques to make 24 hours ahead load forecast for the Power TAC simulation. They found that hour of week, weather related features such as temperature cloud cover were influential to the electricity load demand. The forecasting modules made low error while forecasting for the customers that showed regularity in their energy consumption behavior. The application of linear regression [17] hernandez2012classification and Kalman Filtering [1] also appeared for demand forecasting in the literature.

2.4 Load Forecasting using Clustering

Clustering can be used to group consumers with same electricity demand pattern [8]. The authors [17], applied clustering on 6 months electricity usage of household consumers in

Ireland. Application of clustering generated common load patterns called load profiles which were used to forecast about future load demand. The authors [4] noticed that customers can be categorized to improve accuracy of demand forecasting. They manually clustered the customers in four groups namely, commercial, office, residential and industrial customers. For Power TAC environment, the authors [30] proposed a broker that clusters the bootstrap data of customers and generates a linear regression classifier for demand prediciton for each cluster. The authors [8] used kMeans and Self Organaizing Map to cluster industrial park's consumption in Spain to understand micro environments present in a larger environment.

2.5 Expert System based Load Forecasting

Expert system based electricity demand prediction contains variables that are likely to affect electricity demand [24], [9]. This system then mimicks a human operator's steps to load forecast. This load forecasting mechanism appeared applicable for short term load forecasting [24], [9], [18].

From the review of the literature, the importance of weather related variables such as temperature, cloud cover and wind speed is evident. Also, the hour of the day and day of the week are highly correlated with the load demand. A combination of machine learning classifiers and clustering algorithms appears to be a better idea. For the methodology of [21] it will take a large number of predictors for the simulation system. Also, those predictors will not work if the name of the customer is changed or a new customer is introduced as each predictor is hard coded with a specific customer. It sounds reasonable to cluster the data first and then train machine learning classifier for each cluster. This approach will hold generality. Instead of training only on bootstrap data as proposed in [30], a wealth of data generated from the simulations can be used to train the cluster. Since the clustering is done offline, the proposed approach will not suffer from the problem of having a time limit that the broker has to face if the cluster is trained during the competition. After the clustering is done, for each cluster, different machine learning classifiers can be trained to figure out which one performs the best. So, the broker will no longer stick to linear regression as in [30].

Chapter 3

Customer Description

In this chapter, I will describe the customers present in the PowerTAC simulation system and some statistics their attributes.

3.1 Customer Categories

In Power TAC simulation, a customer can be electricity consumer or producer based on the power type it has. A customer evaluates the tariff plans targetted for its power type and can look for the tariff that gives it maximum monetary benefit. There are several types of customers in the Power TAC simulation such as consumption, interruptible consumption, thermal storage, solar production, wind production and electric vehicle. Each power type has its own characterstics. For example, interruptile consumption customers can shift their electricity demand to some off peak hour, the solar production customers can produce energy based on the weather condition. As opposed to previous methods on demand forecasting, I argue that each category of customers based on the power type should be treated differently. One load forecasting method can be suitable for a category of customers while it may be unsuitable for other categories because each category behaves differently. In the below section, I describe the characterstics of the customers.

- **Consumption:** A customer with power type consumption are the most common customers. They use the energy when they need it. They cannot shift their demand to a future timeslot. Usually, they have a regular pattern of their energy usage. Often, they show a similar pattern for weekdays. They have similar kind of usage pattern for the weekends.

- **Interruptible Consumption:** Interruptible customers are smart enough to shift their energy demand in a timeslot where they can buy electricity at a reduced price. Because of this shifting capability, they don't show a regular usage pattern as the consumption customers do.

- **Thermal Storage:** Thermal storage customers show a weekly pattern in their electricity usage. Also, during a day, their electricity usage in a day depends much on the energy they used in the last timeslot.

- **Solar Production** The solar energy production customer's energy production depends on the cloud cover. They are highly likely to produce energy during the day time.

- **Wind Production** Wind production customers generate energy from the wind.

- **Electric Vehicle** An electric vehicle customer represents one electric vehicle. Their usage of energy is quite irregular and hard to predict.

From figure 3.1 to 3.6 figure shows zScore value of in different hours of week slot. The zScores give a sense if there is a existing pattern in electricity usage. Looking for pattern was useful in designing demand forecasting mechanism for each category of customers. To discuss behavior of customers, a representative customer was chosen. I chose a representative customer from each customer. The following table shows the customer chosen to discuss -

Power Type	Customer Name
Consumption	downtown offices
Interruptible Consumption	village 1 ns
Thermal Storage	sf2
Solar Production	sunnyHill
Wind Production	windmill 1
Electric Vehicle	high income 1

3.2   Statistics

In this section, I present some statistics on the customers available in the system.

- **Customer Vs PowerType**

  In the figure 3.7 we can see the system has more customer with the power type electric vehicle than any other power types. This is because the electric vehicle represents a population of size 1.

- **Population Vs PowerType** From figure 3.8 by far the powertype of consumption has the most number of population.

- **Total Energy Consumed Vs PowerType** From figure 3.9 we can see that the consumption type customers uses the most amount of the electricity.
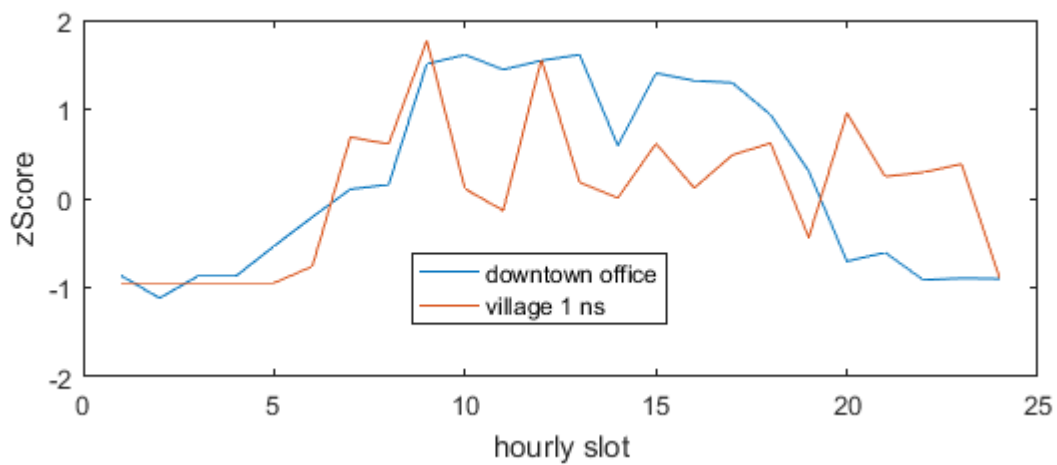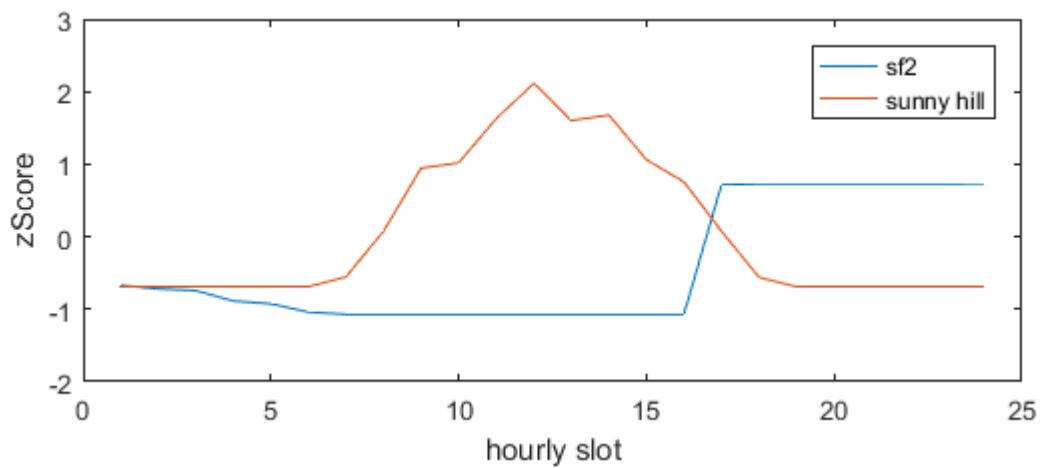
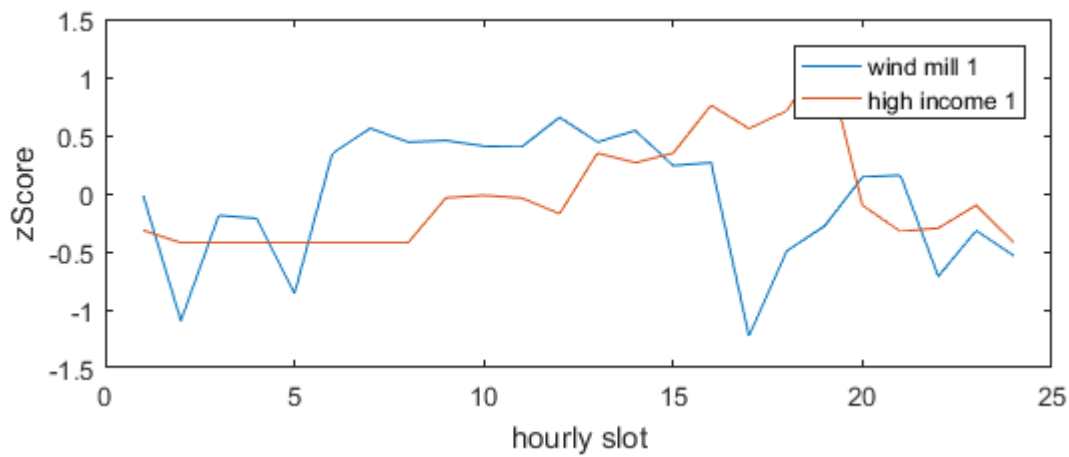Figure 3.1: Caption 1

Figure 3.2: Caption 1
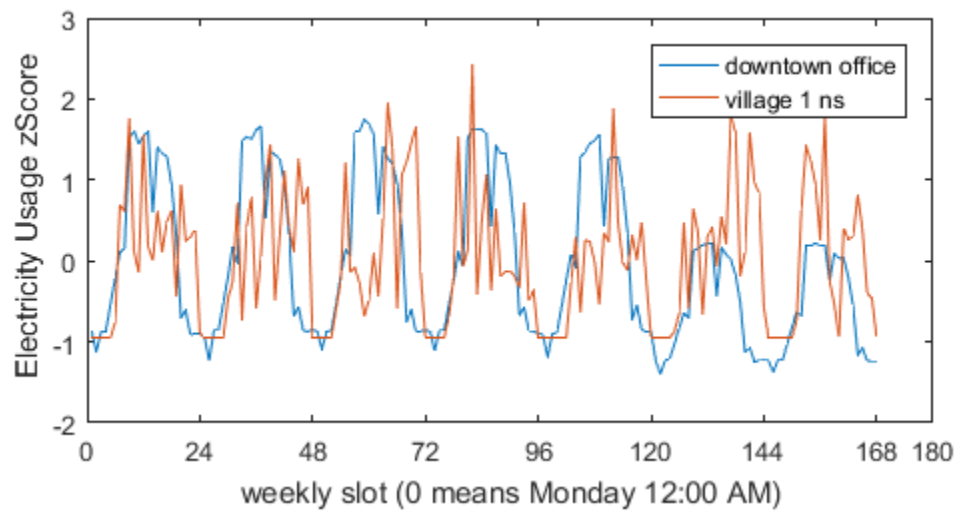

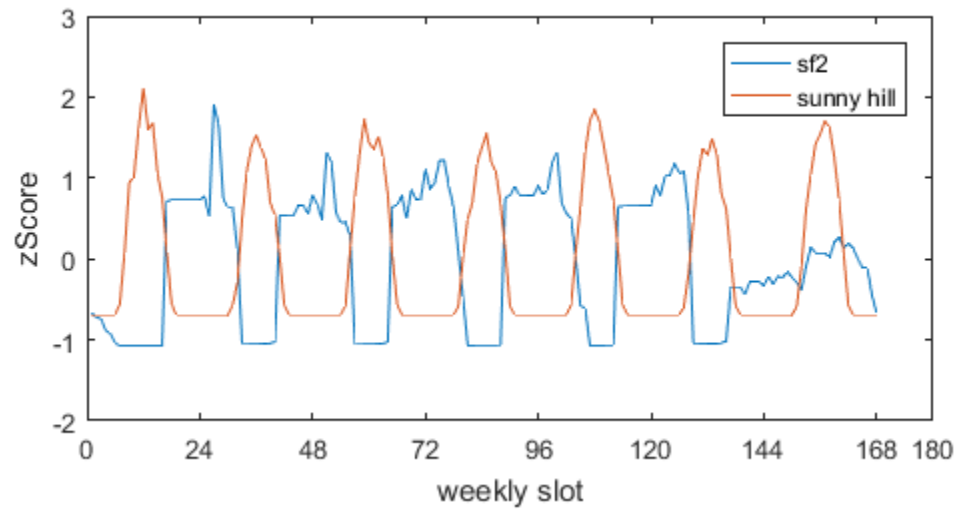Figure 3.3: Caption 1

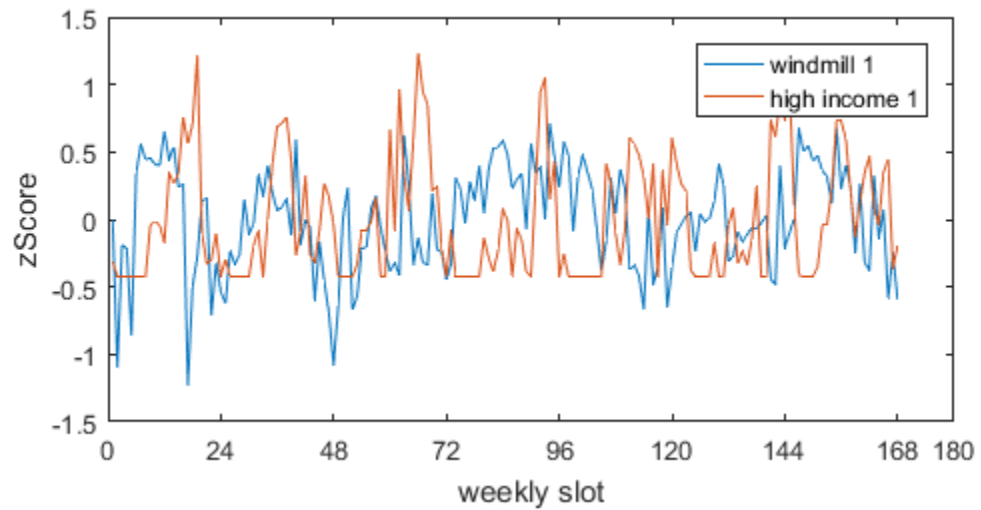Figure 3.4: Caption 1

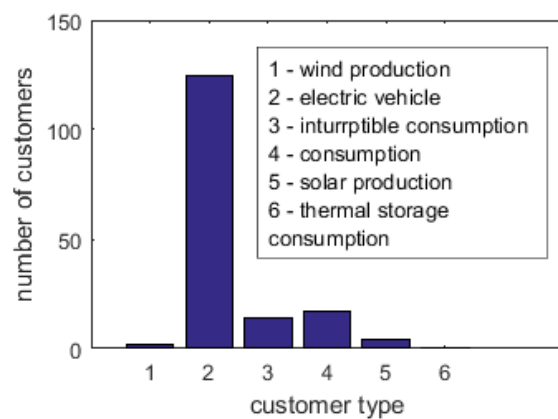Figure 3.5: Caption 1

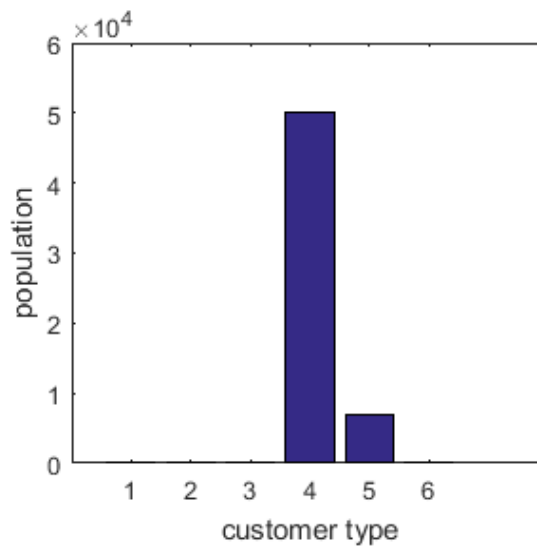Figure 3.6: Caption 1

Figure 3.7:
Number of customers vs Powertype.
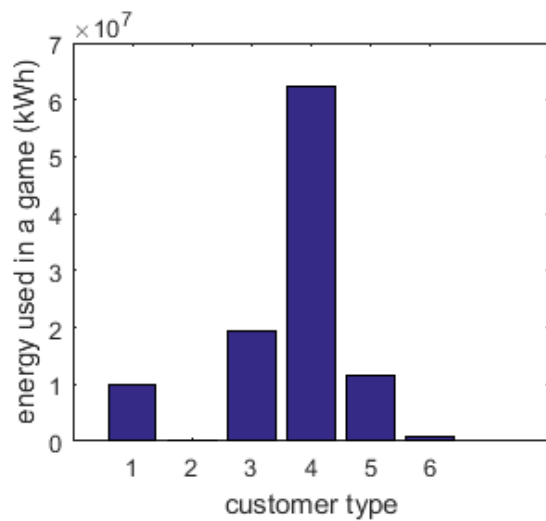

Figure 3.8: Population vs Powertype
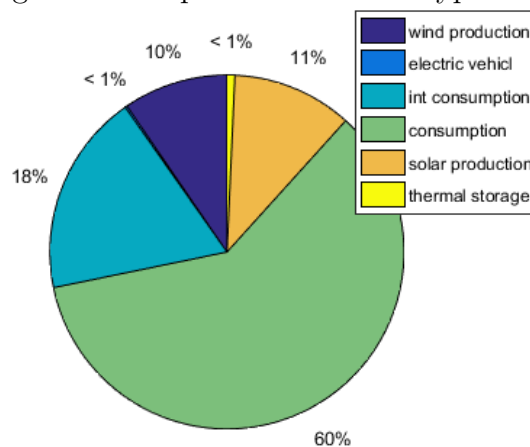

Figure 3.9: Energy vs PowerType.

Figure 3.10:
Energy share for each power type.

Chapter 4

Methodology and Result

Traditionally, a single type of predictor served to predict the energy demand of all power type customers. Since each power type customers acts differently, I have attempted to attack each type of customer separately to make a prediction mechanism that performs better than the baseline predictor.

4.1 The Baseline Electricity Forecasting Mechanisms

The first baseline energy forecasting mechanism is the default prediction mechanism provided by the PowerTAC system. It exploits the fact that usage of a timeslot of a customer in a specific date is highly correlated with the day of the week and the time slot. To make a prediction it stores the average energy usage of an hour of a week. So, for each customer, it uses $24 * 7 = 168$ memory to remember average usages. As soon as it learns about a new usage information of an hour of a week, it updates old average using the following algorithm.

Algorithm 1 Update average usage for $customer_i$ for day d and timeslot t, $newUsage$

1: avgUsage = get average usage of $customer_i$ at day d and time slot t
2: $avgUsage = 0.7 * avgUsage + 0.3 * newUsage$

Algorithm 2 forecast usage for day d and timeslot t for $customer_i$

1: avgUsage = get average usage of $customer_i$ at day d and time slot t
2: return $avgUsage$

The second baseline forecasting mechanism is designed to make energy forecasts for a single customer. In general, if there are n customers in the system, we will need n energy forecasters each one trained on the data of a single customer. I went further by checking different machine learning algorithms such as M5Tree, Linear Regression, M5P rules and REP tree for each customer and picked the best performing one for each customer.

4.2 Proposed Electricity Demand Forecasting Mechanism

In this section, I will describe how I attempted to make energy demand forecaster for consumption power type customers.

4.2.1 Demand Forecasting for Consumption Type Customer

For the consumption type customers, the algorithm 3 describes the proposed method of forecasting energy demand and how it was compared to the baseline methods.

Algorithm 3 Make electricity demand forecasting for consumption type customer

---
1: Extract features for each time slot for each customer [**algorithm 4, 5 and 6**]
2: train kmeans cluster for different sizes of k [**algorithm 7**]
3: train linear regression classifier for each cluster and compute error [**algorithm 8**]
4: pick suitable value for k by observing the errors
5: for each cluster, find the best performing predictor for that cluster [**algorithm 9**]
6: train individual classifer for each customer to make the second baseline [**algorithm 10**]
7: evaluate performance using test data [**algorithm 11**]

The algorithm 3 begins with extracting information from the game log files. All the activities that occurred in a game can be found in a game log. In power TAC the shortest time unit is an hour and it is called time slot. Activities such as buying or selling electricity occur during a time slot. At the beginning of a time slot, the system notifies the broker that a new time slot is about to begin. The system also notifies the brokers with weather forecast about the future time slots. As a time slot ends, the broker receives information about its customer's energy usage which is called tariff transaction report. Algorithm 4 refers how the extraction program retrieves necessary information from tariff transaction report. As the broker gets notification of the beginning of a new time slot, the extraction program has all the information related to energy usage and weather data of the previous time slot available by this time. Algorithm 5 shows the procedure of writing the information of the known time slot's information in training instance file. Once the simulation ends, the extraction program knows the average energy usage of all the customers during a week. In a week there are 24 * 7 = 168 hours or time slots. The extraction program writes all the 168 hourly averages of a week to a file. This is explained in algorithm 6.

Algorithm 4 extract information from transactionReport sent to broker after each time slot through TariffTransactionHandler call back method

---
1: timeSlot = get time slot from transactionReport
2: customerName = get customer name from transactionReport
3: energyUsed = get energy used from trom transactionReport
4: addUsage(customerName, timeSlot, energyUsed)

Next, all the average weekly usages are combined together to make training set for the clustering algorithm. I have used kMeans clustering algorithm to cluster the training set. I have trained clusters of size = 4, 5, 6, 7, 8, 9, 10 and 11. The algorithm 7 describes the procedure of making clusters from the training instances. Once a kMeans of cluster size k is made, a program groups the hourly usages of the customers in the same cluster and combines them to make training set for machine learning classifier. This training set is

Algorithm 5 write extracted data after timeSlot update message received from TimeSlotUpdateHandler call back method

1: knownTimeSlot = timeSlot - 1
2: **for** each customer **do**
3:    day = get day of knownTimeSlot
4:    hour = get hour of knownTimeSlot
5:    statisticalData = get statistics of the customer of day and hour
6:    weatherData = get weather data of knownTimeSlot
7:    trueUsage = get true usage of customer in knownTimeSlot
8:    trainingInstance = create training instance by combining statistical data, weather data and true usage
9:    writeToFile(trainingInstance)
10: **end for**

Algorithm 6 write average electricity usage of the customers of each hour of the week

Require: information of all timeslots has been received
1: **for** each customer **do**
2:    trainingInstance = create empty training instance
3:    **for** each day of week **do**
4:       **for** each hour of day **do**
5:          averageUsage = get average usage of day and hour of customer
6:          append averageUsage to the trainingInstance
7:       **end for**
8:    **end for**
9:    writeToAvgUsageFile(trainingInstance)
10: **end for**

---

used to train linear regression classifier. To test the performance of the classifiers, I have separated five game logs and they were not used for training purposes. The algorithm 8 describes how the cluster based predictor's performance was evaluated.

Algorithm 7 create kmeans cluster of size k from weekly usage training instance file

1: data = load weekly average usage file
2: kmeansCluster = build kmeans cluster of size k
3: save kmeansCluster

Based on the errors observed from different kMeansCluster based predictions, I fixed the number of clusters. Once the number of the clusters was fixed, a program creates several machine learning predictors to see which one performs best for a given cluster. The machine learning classifiers that were tried out are linear regression, M5P rules, M5 Tree, REP tree. In the runtime, a customer will be grouped in a cluster based on its weekly usage. Once the program knows the cluster assigned to a customer, the program will load the corresponding demand forecaster to make electricity demand forecast about the customers.

At this phase, I have the proposed cluster-based customer's demand forecaster. Next,

Algorithm 8 find error of kmeans clusters of different size

1: **for** each cluster size k **do**
2:    get the kMeansCluster of size k
3:    **for** cluster in KMeansCluster **do**
4:       combine slot based training instances of that cluster
5:       train linear regression classifier based on the combined data
6:       save the classifier for cluster
7:    **end for**
8: **end for**
9: **for** each training instance **do**
10:    compute error of the instance using each kMeansCluster
11: **end for**

Algorithm 9 find best classifiers of each cluster of kmeans cluster of size k

1: **for** each cluster in kMeansCluster **do**
2:    combine slot based data of the all the customers in cluster
3:    train available classifiers on the combined data using 10 fold cross validation
4:    choose the classifier with minimum error
5:    save the classifier for making demand forecasting for cluster
6: **end for**

---

the baseline predictor that needs a machine learning classifier for each customer is built. At first, the training instances are combined based on the name of the customer. This means for n customers n training set is constructed, each of the training set has only the information of a single customer. A training set related to a customer is used to create machine learning classifiers for that customer. Several classifiers had been tried out to figure out which classifier performs the best for a customer. The best performing classifier was chosen to predict about a customer. Algorithm 10 explains the procedure of getting the best classifier.

Algorithm 10 find best classifiers created for each individual customer

1: **for** each customer **do**
2:    combine all slot based training instance of the customer
3:    train available classifiers on the combined data using 10 fold cross validation
4:    choose the classifier with minimum error
5:    save the classifier for making prediction about the customer
6: **end for**

The next phase is testing the performance of the proposed and baseline methods. For testing, I had used five game logs that were not used for training purposes. For each test instance, all three methods output was observed to figure out the performance. The algorithm 11 shows the mechanism of testing.

Algorithm 11 performance evalulation of each method

1: **for** each test instance **do**
2:    classify the test instance using moving average usage [**algorithm 2**]
3:    classify the test instance using individual prediction mechanism
4:    classify the test instance using cluster based predictor
5:    calculate and accumulate errors of each mechanism [**algorithm12**]
6:    update moving average baseline predictor based on the information from the test instance [**algorithm 1**]
7: **end for**
8: find average error from the accumulated errors for each forecasting mechanism
---

Algorithm 12 calculate error from the predicted value and the true value
---
1: absoluteError = abs(predictedValue - trueValue)
2: relativeAbsoluteError = (absoluteError / trueValue) * 100 %

4.3 Result

4.3.1 Finding number of clusters

At first, I have segmented the customer using KMeans clustering algorithm with cluster sizes = 4, 5, 6, 7, 8, 9, 10 and 11. For KMeans with size k, we will have k clusters. For each of the k clusters, I had a linear regression predictor. I observed the relative percentage error and absolute average the above cluster sizes. Figure 4.1 shows the result. From, the figure it is clear that the size of the cluster does not have a big impact on the prediction performance.

To keep things simple, I have decided to choose Kmeans cluster of size 4. When k = 4 was chosen, table 4.1 shows the cluster assignment for each customer. It can be seen that, cluster-0 held most of the offices, cluster 2 held most of the village types, cluster 3 held the medical center, cluster 1 held large housing such as brooksidehomes, centerville homes and large offices such as downtown offices and centerville offices.

4.3.2 Finding best predictor for each cluster

I have used the following features for a given timeslot to train prediction models.

- Temperature

- Cloud Cover

- Wind Speed
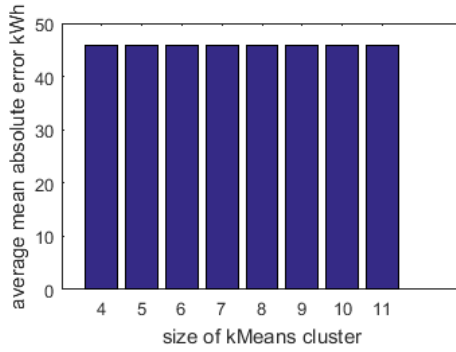
- Average of the Slot

- Standard Deviation of the Slot

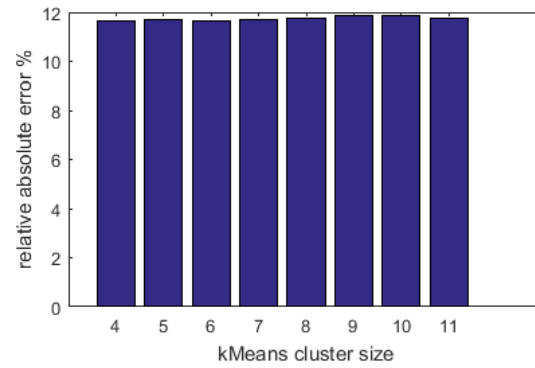Figure 4.1:
cluster type vs average absolute error.

Figure 4.2:
cluster type vs percent relative absolute error

Customer Name	Assigned Cluster Number
BrooksideHomes	0
CentervilleHomes	0
DowntownOffices	1
EastsideOffices	1
OfficeComplex 1 NS Base	0
OfficeComplex 1 SS Base	0
OfficeComplex 2 NS Base	0
OfficeComplex 2 SS Base	0
Village 1 NS Base	2
Village 1 RaS Base	2
Village 1 ReS Base	2
Village 1 SS Base	2
Village 2 NS Base	2
Village 2 RaS Base	2
Village 2 ReS Base	2
Village 2 SS Base	2
MedicalCenter@1	3

Table 4.1: Assigned cluster for each customer

Next, I have tried out M5Tree, Linear Regression, M5P rules and REP tree machine learning classifiers to see which one performs the best for each of the 4 clusters. Figure 4.3, 4.5, 4.7, 4.9 show that M5P, M5P, REPTree and M5RULES are the best predictors for cluster 0, 1, 2 and 3 respectively.

The next step is to find the best classifiers for each of the customers. Based on the data from each of the customers, the four types of classifiers described in previously were tried out. For each customer, the following classifiers performed the best.

The figure 4.11 shows error percentage of each of the predictors type for each of the
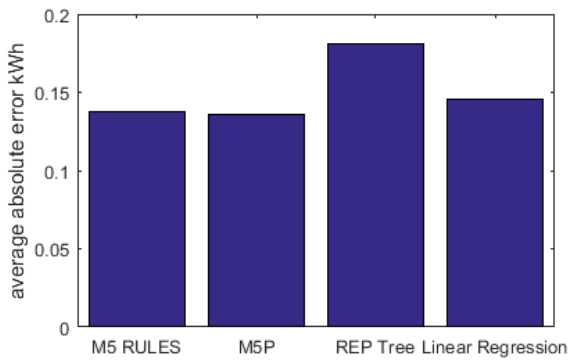
Figure 4.3:
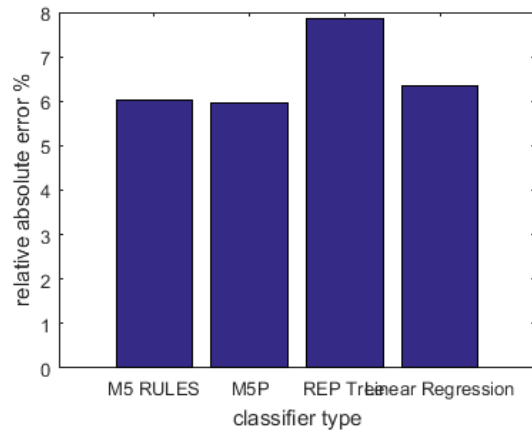cluster 0 average absolute error of 4 classifiers

Figure 4.4:
cluster 0 average relative absolute error of 4 classifiers
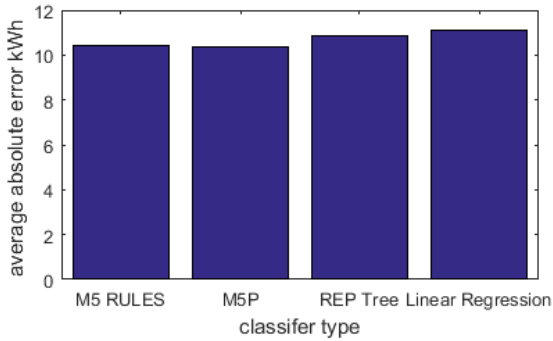

Figure 4.5:
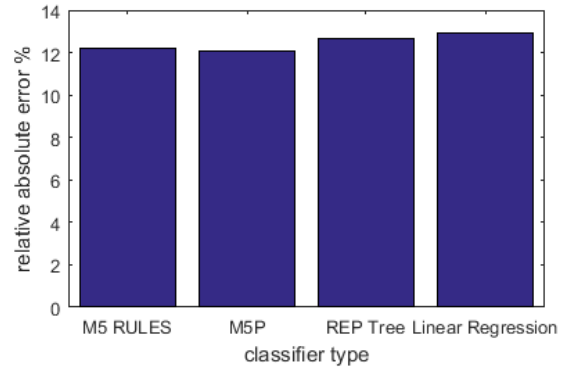cluster 1 average absolute error of 4 classifiers

Figure 4.6:
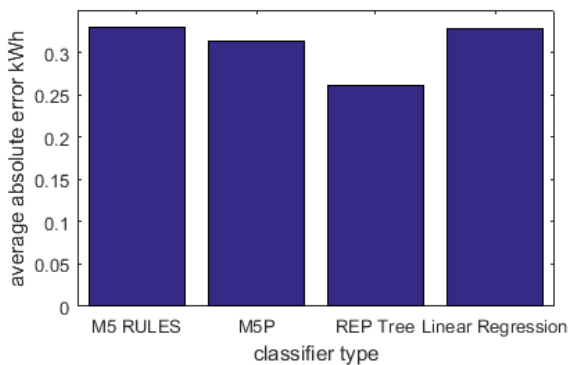cluster 1 average relative absolute error of 4 classifiers

Figure 4.7:
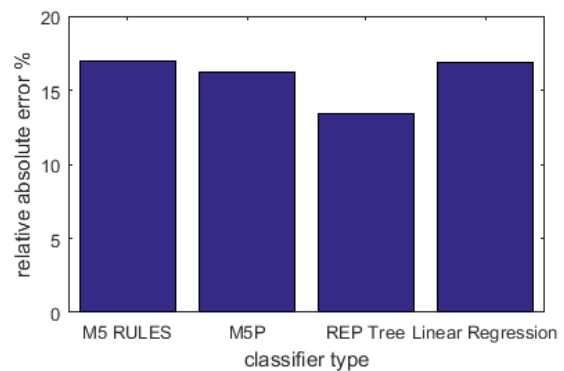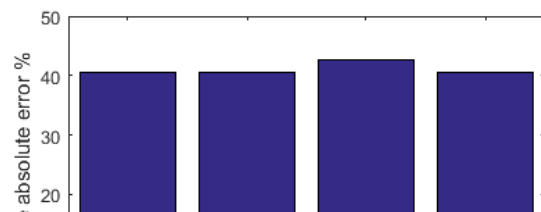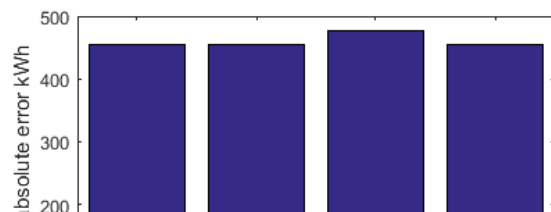cluster 2 average absolute error of 4 classifiers

Figure 4.8:
cluster 2 average relative absolute error of 4 classifiers

Customer Name	Best Predictor Type
BrooksideHomes	M5P
CentervilleHomes	M5P
DowntownOffices	M5P
EastsideOffices	M5P
OfficeComplex 1 NS Base	LinearRegression
OfficeComplex 1 SS Base	LinearRegression
OfficeComplex 2 NS Base	LinearRegression
OfficeComplex 2 SS Base	LinearRegression
Village 1 NS Base	M5P
Village 1 RaS Base	LinearRegression
Village 1 ReS Base	M5P
Village 1 SS Base	M5P
Village 2 NS Base	LinearRegression
Village 2 RaS Base	M5P
Village 2 ReS Base	M5P
Village 2 SS Base	M5P
MedicalCenter@1	M5P

Table 4.2: Best individual predictor for each customer

customer types.

Finally, the cluster based forecasting and the two baselines were tested with data extracted from 5 test files that were not used for training. From Figure 4.12 we can see that cluster based prediction mechanism performed almost as good as the mechanism where n predictors are needed for n customers. And it did well than the default moving average prediction scheme.
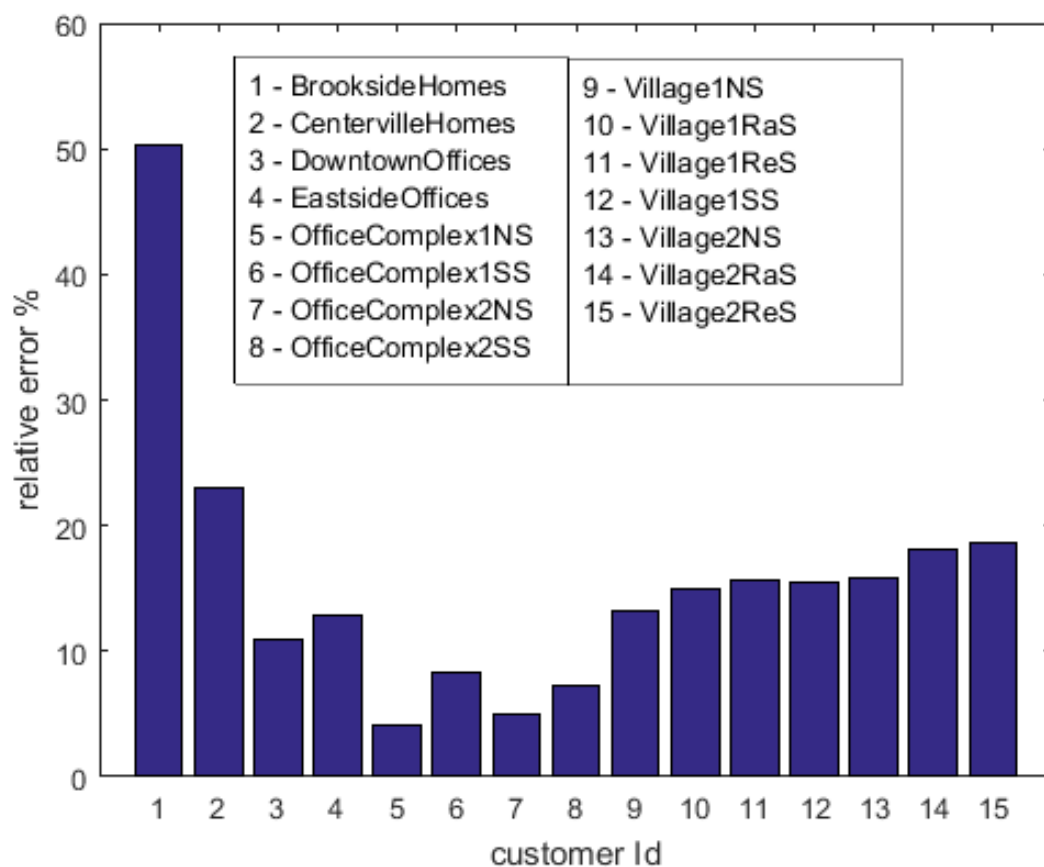
Figure 4.11: Performance of the best classifier for each customer type. Customer Medical center was excluded as it was showing huge error.
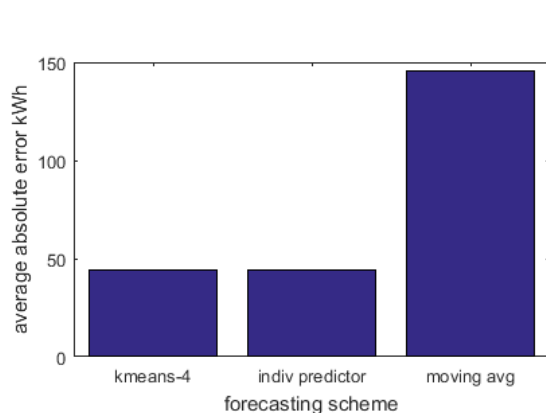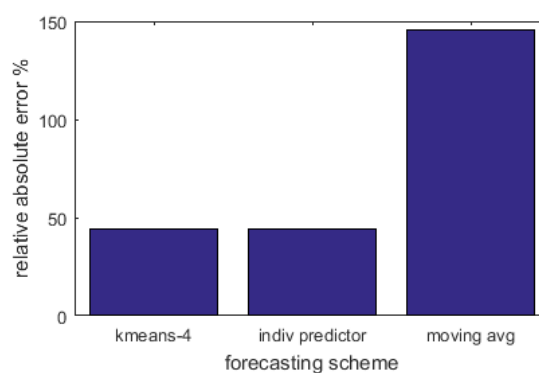

Figure 4.12: average absolute error

Figure 4.13: average percent relative absolute error

References

[1] HM Al-Hamadi and SA Soliman. Short-term electric load forecasting based on kalman filtering algorithm with moving window weather and load model. *Electric power systems research*, 68(1):47–59, 2004.

[2] Nima Amjady. Short-term hourly load forecasting using time-series modeling with peak load estimation capability. *IEEE Transactions on Power Systems*, 16(3):498–505, 2001.

[3] Bo-Juen Chen, Ming-Wei Chang, and Chih-Jen Lin. Load forecasting using support vector machines: A study on eunite competition 2001. *Power Systems, IEEE Transactions on*, 19(4):1821–1830, 2004.

[4] MY Cho, JC Hwang, and CS Chen. Customer short term load forecasting by using arima transfer function model. In *Energy Management and Power Delivery, 1995. Proceedings of EMPD'95., 1995 International Conference on*, volume 1, pages 317–322. IEEE, 1995.

[5] Xi Fang, Satyajayant Misra, Guoliang Xue, and Dejun Yang. Smart gridthe new and improved power grid: A survey. *Communications Surveys & Tutorials, IEEE*, 14(4):944–980, 2012.

[6] Heiko Hahn, Silja Meyer-Nieberg, and Stefan Pickl. Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research*, 199(3):902–907, 2009.

[7] Melissa Hart and Richard de Dear. Weather sensitivity in household appliance energy end-use. *Energy and Buildings*, 36(2):161–174, 2004.

[8] Luis Hernández, Carlos Baladrón, Javier M Aguiar, Belén Carro, and Antonio Sánchez-Esguevillas. Classification and clustering of electricity demand patterns in industrial parks. *Energies*, 5(12):5215–5228, 2012.

[9] Ku-Long Ho, Yuan-Yih Hsu, Chuan-Fu Chen, Tzong-En Lee, Chih-Chien Liang, Tsau-Shin Lai, and Kung-Keng Chen. Short term load forecasting of taiwan power system using a knowledge-based expert system. *IEEE Transactions on Power Systems*, 5(4):1214–1221, 1990.

[10] Ching-Lai Hor, Simon J Watson, and Shanti Majithia. Analyzing the impact of weather variables on monthly electricity demand. *IEEE transactions on power systems*, 20(4):2078–2085, 2005.

[11] Che-Chiang Hsu and Chia-Yon Chen. Regional load forecasting in taiwan—-applications of artificial neural networks. *Energy conversion and Management*, 44(12):1941–1949, 2003.

[12] Ercan Izgi, Ahmet Öztopal, Bihter Yerli, Mustafa Kemal Kaymak, and Ahmet Duran Şahin. Short–mid-term solar power prediction by using artificial neural networks. *Solar Energy*, 86(2):725–733, 2012.

[13] Nur Adilah Abd Jalil, Maizah Hura Ahmad, and Norizan Mohamed. Electricity load demand forecasting using exponential smoothing methods. *World Applied Sciences Journal*, 22(11):1540–1543, 2013.

[14] Wolfgang Ketter, John Collins, and Prashant Reddy. Power tac: A competitive economic simulation of the smart grid. *Energy Economics*, 39:262–270, 2013.

[15] Wolfgang Ketter, John Collins, and Mathijs De Weerdt. The 2016 power trading agent competition. *ERIM Report Series Reference*, 2016.

[16] Bart Liefers, Jasper Hoogland, and Han La Poutré. A successful broker agent for power tac. In *Agent-Mediated Electronic Commerce. Designing Trading Strategies and Mechanisms for Electronic Markets*, pages 99–113. Springer, 2014.

[17] Fintan McLoughlin, Aidan Duffy, and Michael Conlon. A clustering approach to domestic electricity load profile characterisation using smart metering data. *Applied energy*, 141:190–199, 2015.

[18] Ibrahim Moghram and Saifur Rahman. Analysis and evaluation of five short-term load forecasting techniques. *IEEE Transactions on power systems*, 4(4):1484–1491, 1989.

[19] S Ozdemir and R Unland. Agentude: The success story of the power tac 2014s champion. In *AAMAS Workshop on Agent-Mediated Electronic Commerce and Trading Agents Design and Analysis (AMEC/TADA 2015)*, 2015.

[20] Serkan Ozdemir and Rainer Unland. A winner agent in a smart grid simulation platform. In *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, volume 2, pages 206–213. IEEE, 2015.

[21] Jaime Parra Jr and Christopher Kiekintveld. Initial exploration of machine learning to predict customer demand in an energy market simulation. In *Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence*, 2013.

[22] Cameron W Potter, Allison Archambault, and Kenneth Westrick. Building a smarter smart grid through better renewable energy information. In *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES*, pages 1–5. IEEE, 2009.

[23] Hao Quan, Dipti Srinivasan, and Abbas Khosravi. Short-term load and wind power forecasting using neural network-based prediction intervals. *IEEE transactions on neural networks and learning systems*, 25(2):303–315, 2014.

[24] Saifur Rahman and Rahul Bhatnagar. An expert system based algorithm for short term load forecast. *Power Systems, IEEE Transactions on*, 3(2):392–399, 1988.

[25] Andre Richter, Erwin van der Laan, Wolfgang Ketter, and Konstantina Valogianni. Transitioning from the traditional to the smart grid: Lessons learned from closed-loop supply chains. In *Smart Grid Technology, Economics and Policies (SG-TEP), 2012 International Conference on*, pages 1–7. IEEE, 2012.

[26] Ina Rüdenauer and Carl-Otto Gensch. Energy demand of tumble driers with respect to differences in technology and ambient conditions. *Öko-institut, Freiburg*, 2004.

[27] Nicholas I Sapankevych and Ravi Sankar. Time series prediction using support vector machines: a survey. *IEEE Computational Intelligence Magazine*, 4(2):24–38, 2009.

[28] Jonathan Serrano, Enrique Munoz de Cote, and Ansel Y Rodríguez. Fixing energy tariff prices through reinforcement learning.

[29] Daniel Urieli and Peter Stone. Tactex'13: a champion adaptive power trading agent. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 1447–1448. International Foundation for Autonomous Agents and Multiagent Systems, 2014.

[30] Xishun Wang, Minjie Zhang, Fenghui Ren, and Takayuki Ito. Gongbroker: A broker model for power trading in smart grid markets. In *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, volume 2, pages 21–24. IEEE, 2015.