Improved Heat Demand Prediction of Individual Households *

Vincent Bakker * Maurice G.C. Bosman * Albert Molderink * Johann L. Hurink * Gerard J.M. Smit *

* University of Twente, Department of EEMCS, P.O. Box 217, 7500 AE Enschede, The Netherlands (e-mail:v.bakker@utwente.nl)

Abstract: One of the options to increase the energy efficiency of current electricity network is the use of a Virtual Power Plant. By using multiple small (micro)generators distributed over the country, electricity can be produced more efficiently since these small generators are more efficient and located where the energy is needed. In this paper we focus on micro Combined Heat and Power generators. For such generators, the production capacity is determined and limited by the heat demand. To keep the global electricity network stable, information about the production capacity of the heat-driven generators is required in advance. In this paper we present methods to perform heat demand prediction of individual households based on neural network techniques. Using different input sets and a so called sliding window, the quality of the predictions can be improved significantly. Simulations show that these improvements have a positive impact on controlling the distributed microgenerators.

Keywords: Prediction methods, Energy management, Neural network applications

1. INTRODUCTION

Increasing the overall energy efficiency of the electricity network is an important topic in order to reduce the CO₂ footprint. In recent years, a lot of research and development has taken place to improve the efficiency of electricity systems. Besides energy savings by decreasing the energy demand, the production of electricity can also be improved significantly.

Traditionally, most western countries supply domestic electricity demand through generation in large central power stations, with subsequent transmission and distribution through networks. The generation efficiency of the power stations varies between around 35% (older coal stations) to over 50% (modern combined cycle stations), averaging to about 39%. When transmission and distribution losses are considered, the average overall efficiency of the system drops to 35% (de Jong et al., 2006).

In the coming decade a strong trend towards distributed electricity generation (micro-generation e.g. solar cells, micro Combined Heat and Power (microCHP) appliances, micro gas turbines, micro windmills, heat exchangers, etc.) is expected. In this paper, we focus on microCHP appliances. However, the results are generally applicable. MicroCHP appliances are systems that produce heat and — as a by-product during the heat production — electricity. Current generation microCHP appliances are fuelled by natural gas. They can generate electricity at the kilowatt level which allows these units to be installed in an individual home. They are connected directly to the domestic heating and electrical systems, which leads to a very high efficiency (up to 97%) in usage of primary energy. The heat is used for the heat demand in the home such as central heating, showering, hot water taps etc. The electricity can be used in the home or, when not needed, be exported to the electricity distribution network.

The (electricity) production of a microCHP appliance is heat driven, since it only produces electricity while producing heat. Adding a heat buffer (hot water tank) decouples the demand and production of heat, within the limits of the heat demand and the buffer size. This gives some flexibility in the electricity production, allowing the production of electricity on more beneficial periods. For example, a peak in the electricity demand can be seen when people get home. During this period, electricity can be generated by the microCHP system and used within the home by the appliances switched on. The heat can be used for central heating or to fill the hot water tank. The stored heat can be used the next morning for showering.

It is expected that microCHP appliances will replace the current high efficiency boilers (United States Department of Energy, 2003). This will increase the amount of microCHP appliances on the grid in the near future substantially. When the number of microCHP appliances becomes high enough, generators can be virtually grouped together and become a Virtual Power Plant (VPP). By controlling and smart scheduling such a fleet of generators a VPP may replace a conventional (less-efficient) power plant. Using a VPP instead of a conventional one will result in a significant reduction in costs and CO₂ emission due to a better use of primary energy sources.

Important in the above mentioned VPP approach is the controllability of the large group of generators. As described in Section 2, an accurate heat demand prediction is essential for good controllability. In this paper we present an improved heat demand prediction model, based on an initial approach described in Bakker et al. (2008a), by either using more detailed data and/or by using a different way of learning.

* This research is funded by Essent and GasTerra
After discussing some related work on load prediction in Section 3, the developed model is explained in Section 4. After giving the approach in Section 5, the results are presented in Section 6. We conclude this paper with conclusions and future work.

2. VIRTUAL POWER PLANT

A VPP using highly efficient micro-generators gives a lot of benefits. First, since the micro-generators have a high efficiency, the overall energy efficiency increases. This results in a reduction of the usage of primary energy sources and a reduction of CO₂ emissions. Furthermore, since the electricity/heat is used where it is produced, there are less transport losses, increasing the overall performance even further.

However, controlling a VPP differs from a regular, conventional power plant. The whole system consists of a very large group of small generators, placed on a geographically wide spread area. Several issues like reliability, scalability and communication now have to be addressed. Furthermore, a conventional power plant is not restricted by heat demand since heat is dumped.

One important factor of the electricity network is stability. Stability is reached by keeping the supply and demand in balance at all times. In case of the Dutch electricity market, suppliers (producers) and consumers (retailers) of electricity have to specify one day in advance what their electricity production/demand is going to be for each quarter of an hour. Every deviation from this specification will result in an imbalance and is penalised by a central authority. The deviation has to be compensated elsewhere in the network.

Since demand and supply always have to be in balance, multiple electricity markets exist. The electricity demand is dependent on seasonal changes and short term (local) environmental influences (for example the local weather or an important soccer match on TV). The ‘base’ load which solely depends on long term changes (like the seasonal changes) is most often traded using long term contracts. The short term fluctuations are traded on a day-ahead market, which trades electricity with a granularity of one hour. Since these short term fluctuations are very dynamic, a production facility which can handle this dynamic behaviour is required. Since this is more difficult to achieve, the prices on the day-ahead market are much higher compared to the long term contracts. The most expensive electricity is the electricity required for the real-time stabilisation of the grid, which is traded per quarter of an hour on very short terms.

Since a VPP consists of many small generators, which can start and stop within a couple minutes, a VPP has the potential to be used on the short-term markets. Dependent on the stakeholder of the VPP, the VPP can be used for balancing the grid by a network company or to reduce purchase costs/penalties by a utility.

Essential in all applications is the reliability of the VPP system. To start off with, the production capacity of the fleet has to be predicted at reasonable accuracy. This ensures that the promised production capacity is really available. Furthermore, communication between all the generators has to be fast and reliable to ensure that proper control is possible.

One of the most important problems is determining the available production capacity and steering the available electricity production. Since a microCHP only generates electricity while producing heat, the production capacity of a fleet of microCHP appliances is dependent on the heat demand of the individual households. The electricity production pattern thus limited by and dependent on the heat demand.

The required production pattern of a VPP is dependent on the objective of stakeholder of the VPP and should be determined centrally. However, centrally deciding which appliance to use when for electricity production is infeasible due to the large size of the production fleet. In Bakker et al. (2008b) a hierarchical scalable system to control a large fleet of generators is proposed. In this approach, a learning system located in each individual household determines the heat demand of the household for the coming day. From this, the production capacity can be deducted and is transmitted to a (central) control system. Based on these predictions, a global schedule is made for the whole fleet. This global schedule is broadcasted to the fleet, and based on this schedule a local schedule for each micro-generator is determined.

In the above mentioned approach, the heat demand for each individual household is predicted using neural network techniques. The goal is to predict the heat profile for the next day as accurately as possible. Bakker et al. (2008a) have shown some promising results, however, for a proper functioning VPP control system, a better prediction of the individual domestic heat demand is needed.

There are several reasons why in Bakker et al. (2008a) and in this research individual heat demand prediction is used. Eventually, the control system has to decide which generators should be switched on or off. Since the fleet consist of hundreds of thousands up to a million of households, it is infeasible to do a prediction per household centrally. By moving the prediction to a local control system in the house, a scalable system is achieved.

Based on the local heat demand prediction a schedule for the generator is made. The heat demand prediction must thus be accurate. Heat demand is dependent on external factors like weather, insulation and human behaviour. Each house is different and has different insulation characteristics. Every household is different and has different behavioural patterns. By predicting the heat demand per household locally, local information about the specific environmental and behavioural characteristics can be learned to improve the prediction.

The prediction model should be able to learn the behaviour of the people. Since people have different behaviour on different days, the model has to be flexible. Changes in behaviour should be learned quickly in order to cope with changes like holidays. Furthermore, it should be able to use information about external factors, like weather, and to learn the (more or less fixed) characteristics of the house.
3. RELATED WORK

The prediction of loads is a well studied topic. As stated above, the demand is determined by a) environmental factors, like weather influences and insulation and b) social behaviour of the consumers.

In Asar and McDonald (1994); Chen et al. (1992) predictions of electricity loads using neural networks are described. As input data historical demand data and multiple combinations of weather influences are used.

In Dotzauer (2002) heat demand prediction for large scale systems is described. The presented model forecasts the load by combining two functions. The first function is a (piecewise) linear function used to determine the influence of the weather and uses outdoor temperatures as input. The second function is used to describe the social factor and is modelled as a constant (total load – weather dependent part). During training, the best combination of the two functions is searched.

In Serban and Popescu (2008) heat demand prediction for district heating systems is described using times series analysis. A time series can be described as output of a system that has as input a white noise signal. The output consists of a (linear) combination of observed data (history) and the present input.

As in Asar and McDonald (1994); Chen et al. (1992), we use neural networks to do load prediction. However, we do not predict electricity loads, but heat loads. Furthermore, we do prediction for households instead of large systems. The prediction for households differentiates this research from Dotzauer (2002); Serban and Popescu (2008), since they do prediction of large scale systems. Prediction for large systems is easier, since differences in behaviour between different (kind of) consumers get smoothened due to the large size of group. Since the schedule of the local generators is dependent on local heat demand, a local heat demand prediction is required in our approach.

4. PREDICTION MODEL

The approach presented in this paper uses neural network techniques for the heat demand prediction. Neural networks are computational models based on biological neurons (Krose and van der Smagt, 1993). They are able to learn, to generalise, or to cluster data. A neural network consists of a pool of simple processing units called neurons, which communicate by sending signals to each other over a large number of weighted connections. A network has to be configured (trained) such that the application of the network to a set of given inputs produces the desired outputs (which are also given). The given input/output pairs form the training data. During training, weights of the connections between the neurons are adapted according to a learning rule. By adjusting the weights, the error between the network output and the desired output is minimised.

If a priori knowledge is available, this can be used to pre-specify the weights.

We use a multi-layer feed-forward network. Since we have no a priori knowledge, our model has to be trained completely. We assume the most relevant factors for the heat demand are the weather, the characteristics of the house and the behaviour of the residents. Since houses do not change that often, we consider the characteristics of the house static. Because of this, the neural network should be able to learn the characteristics since they are present in all data used.

As people behave differently on different days of the week due to e.g. social activities, we use seven neural networks, one for each day of the week. A network has 24 outputs, each output representing the heat demand of one our of the day.

To learn the behaviour of the residents, historical heat demand is used as an input. In order to learn a pattern for a specific day in the week, the heat demand one week earlier is used as input. The heat demand one day earlier is also used to incorporate recent behaviour. To incorporate external factors, outdoor temperatures are used as input.

Although the results using this structure from Bakker et al. (2008a) are promising, a better prediction of the heat demand is needed to improve controllability of the VPP.

In this work, two methods to improve the predictions are described. By adding more input factors or by improving the quality of the input, a decreased prediction error should occur.

As mentioned above, it is assumed that heat demand is dependent on weather, environment and behaviour. Since only the weather and behaviour are dynamic, more information about these factors should be added to the input set.

For the weather factor, we add wind speed information to the input set. Information about temperatures and wind speeds are obtained from weather stations nearby the households. These weather stations output Meteorological Aerodrome Reports (METAR) every half hour, from which the temperatures and wind speeds are extracted.

For the behaviour factor, information about the thermostat program and user overwritten settings might give an improvement in prediction. However, the assumption is that the thermostat program is fixed and can thus be learned by the system. If the thermostat program is not fixed, it is dependent on the behaviour of the residents and cannot be determined on before hand. Other information about the user, for example about holidays, is not used since this requires interaction with the user. We aim for a system running autonomously.

Since there is limited extra information available, the first method to improve the predictions is to add values containing wind speed information to the input set.

The second method to improve the predictions is to improve the quality of the input set. Heat demand data of four households from end 2006 up to end 2007 are used to train and validate the models. In Bakker et al. (2008a), this data set was split up into a training set and a validation set choosing random weeks in the year. Using cross-validation, the performance of the prediction models was determined. The reason for choosing random weeks for training is to make the prediction more general and to find as much behaviour as possible. However, in case of human behaviour, this might not be the best choice.
In this paper we try to improve the quality of the input set, by instead of using random weeks of the whole data set, only the last weeks for a training set. This implies that each day, (re)learning is required using information of the last weeks. By using this ‘sliding window’, more recent behaviour is used to learn the behaviour of the residents. This makes the prediction model more flexible and adaptable. Furthermore, the assumption is that behaviour changes during the year, and that current behaviour resembles more to the recent behaviour instead of general behaviour during the whole year.

One important decision to make while using the sliding window approach is how many weeks of history to use for training. Adding more weeks might give more information about the behaviour, but this can cause to learn ‘too general’ or outdated behaviour. Using too few weeks can cause ‘overfitting’ to the recent behaviour, missing the general behaviour people might have on a certain week day.

5. APPROACH

To analyse the impact of the changes in the input set on the prediction, all combinations of the proposed methods for improvements have been calculated. The reference case is not adding wind speeds and no usage of a sliding window. The reference input set consists of the heat demand one day before the predicted day (24 values), the heat demand one week before the predicted day (24 values) and the outdoor temperatures on the predicted day (assuming perfect weather predictions) (48 values). The reference set contains data from randomly chosen days in the year, given to the neural networks in a random order. The size of the reference set is between 13 and 28 examples (days), dependent on the household.

To analyse the impact of adding wind speeds to the input, the same days of the year are used for training to give a fair comparison. On top of the reference set, 48 values of wind speed information are used during training.

In the sliding window approach, the input sets contain the same input fields as the reference set. The same days as in the validation set of the reference case are predicted. Instead of using a random training set for all days in the validation set, data just before the day to predict are used for training. The size of the training set (how many weeks of data used for training) is called the window size. Window sizes of three up to six weeks are used.

Besides looking at the impact of using or adding wind speeds or sliding window solely, a combination of both techniques is used as well. Since there is less training data used in the sliding window, the quality of the input set is of more significance. By adding wind speed information to the data set, it is expected that the quality of the predictions using sliding windows improves.

When using neural networks, choosing the right network size (number of hidden layers, number of hidden neurons) is usually done via cross validation. Since there is a high correlation between the input and the output, we chose to use only two layers (one hidden layer, one output layer). To determine the best network size, we started with using four hidden neurons and kept adding neurons to the hidden layer until the error on the validation set increased a few times (this was reached at twenty neurons). Since we have data of four households, for each household and each day of the week networks of sizes from four up to twenty are trained. We have chosen to use multiple training sessions to cope with finding local minima while training the neural networks. Concrete, we use three iterations and use the best performing iteration.

As a measure for the performance of a network, a score of a network is calculated. Two factors are important when predicting the heat demand and thus the production capacity of the generator. The shape of the heat profile and the amount of heat for the day determine the quality of the prediction.

The first measure is how good the network is able to learn the specific profile of that day. We determine the ‘heat profile’ for a day by calculating mean heat demand for each hour of the day over a given set. In other words, the mean heat demand between 12 AM and 1 AM, 1 AM and 2 AM etc. is determined. By subtracting the predicted heat profile from the real heat profile, we get a vector of 24 values with the deviation of the profile per hour. The mean squared error of these 24 values is called the ‘profile error’.

The second measure is how good the network is able to predict the amount of heat per day. When for some reason the profile is predicted badly (for example the profile is delayed for one hour), it is still important that the total production capacity, which can be used for generation, is predicted well. For each day in the data set, the total heat demand for the whole day (the sum of the 24 hours) is calculated. The mean squared error between the real and predicted totals is the second error, the ‘totals error’.

To determine the best performing network, both errors are taken into account. First, both errors are normalised between 0 and 1, since they have different orders. Since we think correct shape is more important than a good prediction of the totals, the profile error is given a three times higher weight than the day total error (0.75 and 0.25 respectively). The result after normalisation and weighting gives the score (performance) of a model, where a lower value means a better prediction.

All combinations of network sizes, features in the data sets (wind/no wind) and data set order (normal vs sliding window) are trained and validated for all four households and all seven days of the week with a developed tool using the Fast Artificial Neural Network Library.

6. RESULTS

Using the developed tool, the impact of the changes in training methods can be analysed. As a measure for the impact, the average relative score is used. The relative score is defined as the new score divided by the reference score. Per approach, the relative scores are determined for each day of the week and each household. The averages of the whole week are depicted in Table 1, presenting the average relative score for each approach and household. Using this measure, a lower value means an bigger improvement.
The impact of adding wind speed information compared to the reference set can be seen in the second column of Table 1. As can be seen in the column, a significant improvement is only made for household number three. Although more information about the weather is given to the model, hardly any improvement can be seen in the other three houses. When using machine learning techniques, there always is a trade off between complexity and performance of the models. Adding wind speeds to the input set makes the network more complex, but the improvement in performance is not big enough to justify this extra complexity. Still an improvement is made, so perhaps a more dense representation of the wind speed information might be a good trade-off between the gained performance and complexity.

The average relative scores using the sliding window approach are given in the last block of Table 1. On average, using a sliding window approach improves the scores of the predictions. Furthermore, using more historical data (a larger sliding window) gives a better performance. Thus, using more training examples gives a better result. This is a bit surprising, since using more historical data will result in learning more general behaviour. The input set of our reference contains between 13 up to 28 weeks of training data. It would be interesting to analyse the impact of using an even larger sliding window. Unfortunately, due to some imperfections of our data set, we do not always have more than six consecutive weeks of data.

The results of the combination of using both a sliding window and using wind speed information during training is given in the last columns of the left block in Table 1. Using this approach, still an improvement in the prediction is made. However, the improvement is not as big as compared to the sliding window approach without the wind speed information. Thus even when less history as training data is given to the model, the extra information about the environment does not improve the predictions.

If we look at the best performing networks for each household and day (thus selecting the best approach for each household and each day), we see that only ten of the 28 best networks use the wind speed information.

From these 28 networks, networks with a window size of three weeks is nine times the best network, and networks with a window size of four, five and six weeks six times. Only once the normal training method performs best. Thus, although a larger window sizes has a larger average improvement in prediction, using a window size of only three weeks more often gives the best performing network.

To give some more insights in the improvements and quality on the prediction, some example plots are given in Figures 1, 2 and 3. In Figure 1 the average profile of

<table>
<thead>
<tr>
<th>House</th>
<th>Using windspeeds</th>
<th>No windspeeds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
<td>3w</td>
</tr>
<tr>
<td>1</td>
<td>89</td>
<td>68</td>
</tr>
<tr>
<td>2</td>
<td>99</td>
<td>63</td>
</tr>
<tr>
<td>3</td>
<td>63</td>
<td>86</td>
</tr>
<tr>
<td>4</td>
<td>114</td>
<td>82</td>
</tr>
<tr>
<td>Avg</td>
<td>91</td>
<td>75</td>
</tr>
</tbody>
</table>

Table 1. Average relative scores (in %)

Fig. 1. Heat profiles for household three on Mondays

Fig. 2. Total heat demands for household three on Mondays

Fig. 3. Heat demand for household 3 on 2007-10-29

Fig. 4. Deviation from global schedule using no sliding window (in no. of microCHPs)

Fig. 5. Deviation from global schedule using sliding window (in no. of microCHPs)
household three of Mondays is depicted. In the figure the prediction of the network training using normal training and the prediction of the networks trained using a sliding of six weeks are depicted. In both cases, no windspeed information used. We see that the sliding window approach follows the trend of the heat profile better then the normal approach.

In Figure 2 the total heat demands of the fifteen test days in the validation set are depicted. Here we can see that the sliding window approach is better able to predict the total heat demand per day.

In Figure 3, the heat demand predictions of households three for October 29, 2007 is depicted. For this day, we have simulated our VPP control system for all four houses. Our approach consists of three steps: 1) Prediction (as described in this paper), 2) Global Planning of a fleet and 3) Local Control.

Given the predictions of the four houses, an optimal schedule of the microCHP has been generated. The global scheduler, described in Bosman et al. (2009), uses an Integer Linear Programming solver to find the optimal runtime of the microCHP appliances. As constrains to the solver, the limitations of the microCHP and the heat buffer are used. The optimisation objective is to maximise the profit made by the microCHP, using the electricity market prices.

Using this optimal schedule, these four households with the local controller have been simulated using our simulator (Molderink et al., 2009). The local controller tries to obey to the global schedule, within the real electricity and heat demand of the house. Here, real heat demand is used and the errors in the heat demand prediction are reflected in a deviation in the runtime of the microCHPs. The deviations between the optimal schedules according to the global scheduler, and the real production patterns are depicted in Figures 4 and 5.

Assuming no prediction error, thus using the real heat demand, we still see a deviation between the optimal global schedule and the real production pattern. The model of a microCHP appliance in the global scheduler is less advanced then the model of a microCHP in the simulator. This explains why there is an imbalance of 4 kWh. The sliding window approach shows an imbalance of 15 kWh, the ‘normal’ approach 16 kWh. As can be seen in the figure, the ‘sliding window’ deviations curve is closer to the perfect prediction schedule than the ‘normal’ one.

Looking at the price per kilowatt using the APX prices for this day, the perfect case generates 9.47 €cents/kWh, the normal case 7.37 €cents/kWh and the sliding window 7.43 €cents/kWh. We see that the improvement in the prediction gives a higher price.

7. CONCLUSIONS AND FUTURE WORK

To improve the prediction of the heat demand of individual households, two new approaches are presented. The first one, adding wind speed information to the input data set, has only a small effect on the predictions of the heat demand. The second technique, using a sliding window (the most recent behaviour) instead of using normal, random training data, yields a large improvement. Instead of learning ‘general’ behaviour of the residents during a longer period of time, it is better to use more recent behaviour and thus only recent heat demands as input. However, given the sliding window sizes, it is better to use a larger window and thus more training data. Using both a sliding window and wind speed information does not improve the prediction compared to the sliding window without wind speed information. Simulations have shown that the improvements in prediction quality leads to a better controllable fleet.

In the current input set, quite some detailed weather information is given. In future work, a more dense form like only minimum/maximum temperatures and wind speeds might be sufficient as weather input, simplifying the model. Furthermore, using a larger window size has to be analysed.

REFERENCES


