BayesForSG: A Bayesian Model for Forecasting Thermal Load in Smart Grids

Rohan Nanda, Saguna Saguna, Karan Mitra and Christer Åhlund
Luleå University of Technology
Skellefteå, Sweden
{rohan_ramniwas@hotmail.com; saguna.saguna, karan.mitra, christer.ahlund@ltu.se}

ABSTRACT

Forecasting the thermal load demand for residential buildings assists in optimizing energy production and developing demand response strategies in a smart grid system. However, the presence of a large number of factors such as outdoor temperature, district heating operational parameters, building characteristics and occupant behaviour, make thermal load forecasting a challenging task. This paper presents an efficient model for thermal load forecast in buildings with different variations of heat load consumption across both winter and spring seasons using a Bayesian Network. The model has been validated by utilizing the realistic district heating data of three residential buildings from the district heating grid of the city of Skellefteå, Sweden over a period of four months. The results from our model show that the current heat load consumption and outdoor temperature forecast have the most influence on the heat load forecast. Further, our model outperforms state-of-the-art methods for heat load forecasting by achieving a higher average accuracy of 77.97% by utilizing only 10% of the training data for a forecast horizon of 1 hour.

CCS Concepts

•**Applied computing** → Computers in other domains;

Keywords

smart grid; district heating system; Bayesian network; forecasting; machine learning

1. INTRODUCTION

Buildings account for 40% of the total energy consumption in the European Union [8]. The total heating end use consumption (including space and water heating) contributes to 80% of the total household energy consumption [6]. Therefore, it is essential to focus on reducing and optimizing the heating consumption in buildings to achieve the energy and climate targets of 2020 [8]. Forecasting the heat load consumption in buildings aids in developing demand response schemes to match the ever-changing consumer energy needs. This ensures dynamic energy distribution and energy savings in a smart grid. However, predicting the heat energy consumption in the households remains a challenge because of significant variation in the heat load consumption (at the consumer side) due to the presence of a large number factors influencing the heat load forecast such as outdoor temperature, solar radiation, supply and return temperature, wind speed and wind direction [14]. This, necessitates the development of efficient models that can deal with uncertainty arising from the usage of these parameters for heat load forcast in district heating system (DHS) as part of the smart grid infrastructure.

In this paper, we propose, develop and validate an efficient heat load forecasting model: BayesForSG - a Bayesian model for forecasting the heat load consumption of residential buildings in a smart grid. We study and compare the impact of both district heating operational parameters at the substation (belonging to consumers buildings) side and weather forecast (outdoor temperature) parameters on the heat load forecast. We use realistic data traces belonging to the DHS of a major utility company in Skellefteå, Sweden to validate our proposed model. Our results show that BayesForSG achieves higher average heat load prediction accuracy of 77.97% by utilizing only 10% of the training data for a forecast horizon of 1 hour.

The rest of the paper is organized as follows. In the next section, we discuss the background and related work. Section 3 describes the proposed model in detail. Section 4 discusses the results and analysis. Section 5 presents the conclusion and future work.

2. BACKGROUND AND RELATED WORK

District heating system is an infrastructure comprising of three major components: the heat production system, the distribution network and the heat consumption system. The production system consists of a combined heat and power plant (CHP) or a boiler for heat energy production. The heated water (at supply temperature) from the production system is supplied to residential buildings using a network

SAC 2016,April 04-08, 2016, Pisa, Italy

DOI: http://dx.doi.org/10.1145/2851613.2853127

of insulated pipes. The heat consumption system consists of substations at each building for distributing hot water to different consumers inside the building. The cooled water(return temperature) from the buildings is supplied back to the production side using the distribution network. Figure 1 shows a simple diagram of a DHS.

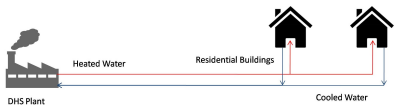

Figure 1: Schematic diagram of a district heating system(DHS).

Since, the objective of this paper is to optimize the district heating grid, by forecasting the heat demand at the consumption side, we explore the relevant related work in this section. The work discussed in [1] uses a wide range of data mining algorithms like random forest, support vector machines (SVM), multi-layer perceptron (MLP), MLP ensemble and k-nearest neigbour (k-NN) to forecast the steam load in a building by utilizing weather forecast parameters such as outdoor air temperature, humidity, solar radiation, barometric pressure, wind speed, rain gauge and wind position. The authors concluded that MLP ensemble method performs best.

The authors in [15] developed an artificial neural network for short-term thermal load forecast in a building complex by utilizing outdoor temperature, pressure and flow rate. The prediction accuracy of this model drops with the increase in forecast horizon. The work presented in [9] utilizes supply temperature, return temperature, flow rate, outdoor temperature, current load and historical load as input parameters for forecasting heat load using support vector regression.

Idowu et al. [5] built a forecasting model to compare the performance of four supervised machine learning algorithms: support vector regression, regression tree, feed forwards neural network and multiple linear regression. They observed that support vector regression achieved the best accuracy. Vlachopoulou et al. [13] used simulated data and expert knowledge to build a dynamic Bayesian network for aggregated water head load consumption in residential buildings. The authors do not provide much clarity about the forecasting accuracy and commonly used metrics for measuring the prediction errors. Also, the model still needs to be validated on the real world data.

Our work differs from related work as we validate our model by utilizing the realistic district heating data from the city of Skellefteå, Sweden over a period of two seasons(winter and spring) in three buildings with different heat load variations. We also evaluate the performance of several parameters on the heat load forecast and identify the most influential parameters.

3. THE PROPOSED MODEL

We now discuss two key requirements that should be satisfied by heat load forecasting models. It is estimated that the collection of energy consumption data through smart meters at a high resolution, could result in a 3000 fold increase in production of data as compared to monthly recordings [10]. This data explosion leads to the problem of analysis of a large amount of data for understanding the user's energy consumption patterns. There are around 5000 substations in the city of Skellefteå and outskirts which generate large amounts of data. Therefore, a key requirement in the domain of thermal load forecast is to choose a machine learning technique which can use the least amount of training data to build accurate and efficient forecast models. Another major requirement focusses on developing a unique machine learning model that can be flexible to different magnitudes of heat load variation across various buildings. The proposed model, BayesForSG validates both these requirements.

3.1 Bayesian Networks

A Bayesian network is a directed acyclic graph that models the probabilistic relationships between the random variables [11]. Each node of the graph represents a random variable. The random variable can either be continuous or discrete. An edge between the nodes is a directed link that connects two nodes. This edge represents the conditional dependency (probabilistically) between the two nodes. We consider a set of random variables $\{x_1, ..., x_n\} \in X$ in a Bayesian network. The joint probability distribution of the network is computed as the product of all the conditional probabilities specified in the Bayesian network as illustrated by the equation below [11].

$$P(x_1, ..., x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i)) \quad (1)$$

Here, $parents(X_i)$ represent the parent node of nodes x_i. Each node in the network has a conditional probability distribution reflected in a conditional probability table.

3.2 Heat Load Consumption Dataset

The Bayesian model developed in this paper, is based on the heat load consumption dataset belonging to the DHS of Skellefteå, Sweden.

Seasons: Heating is primarily used in cold weather conditions when the outdoor temperature is really low. For this reason we choose to develop our model for Winter and Spring seasons. From the dataset available to us, we take the duration from 22 December 2013 to 28 February 2014 for the Winter season. The duration from 1 March 2014 to 30 April 2014 is considered for the Spring season.

Buildings: We consider three residential buildings which have multi-family apartments. We refer to these buildings as Building A, Building B and Building C. The heat load variation in the buildings is shown in Figure 2 and in Tables 1 and 2. The substations at these buildings are deployed with sensors that record the aggregated heat load consumption at an interval of each minute.

Available parameters influencing the heat load: The sensors deployed at the buildings record the district heating operational parameters, which include supply temperature, return temperature and flow rate. The outside temperature is also recorded from the on-site temperature sensor. The

energy meters at the substation calculate the heat load from the DHS operational parameters according to the following equation [12]:

$$Q_{substation} = c * mt' * (T_s t' - T_r t')\, dt' \quad (2)$$

Here $Q_{substation}$ is the heat load consumption at a particular substation. c is the specific heat of the liquid in the district heating distribution network(mostly water), m is the flow rate, T_s is the supply temperature, T_r is the return temperature [12]. From this equation, it can be observed that the heat load consumption also depends on flow rate, supply temperature, return temperature and the difference of supply and return temperature. We refer these parameters as DHS operational parameters. The difference between supply and return temperature is referred as T_{delta}. From the experts in Skellefteå Kraft(energy provider in Skellefteå), we learnt that the heat load demand at the production side is computed by considering the hourly outdoor temperature, T_{out}. Therefore, we consider outdoor temperature as a key parameter influencing the heat load demand. The fact that outdoor temperature varies over different seasons, further motivates the need for a seasonal load profiling model for residential heat load consumption. Figure 2 shows the heat load variation across all three buildings during winter season. Each building has its own trend of variation in heat load. The challenge is to develop a forecasting model that can capture this variation by making accurate forecasts.

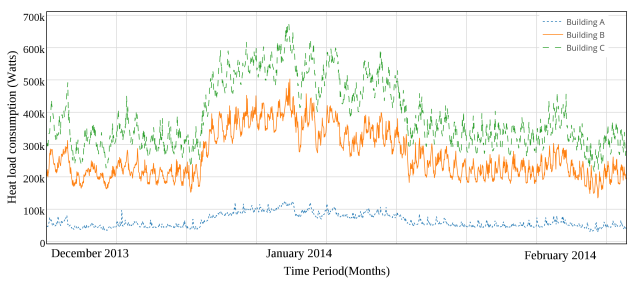

Figure 2: Heat load variation across all three buildings during winter season.

Table 1: Heat load variation in Buildings A, B and C during winter season.

Heat Load	Minimum (Watts)	Maximum (Watts)	Mean (Watts)	Standard Deviation(Watts)
Building A	29385	124533	63703	21124
Building B	134135	503548	267335	73485
Building C	213141	675061	389995	102743

Table 2: Heat load variation in Buildings A, B and C during spring season

Heat Load	Minimum (Watts)	Maximum (Watts)	Mean (Watts)	Standard Deviation (Watts)
Building A	12088	91425	42133	13268
Building B	74438	358408	188629	45278
Building C	135600	482268	283675	62801

Additional parameters: In order to study the effect of user behaviour and daily load patterns, we added hour of day and day of week as two additional parameters influencing the heat load forecast. These parameters are added as discrete variables to the dataset. For example, two values *Weekend and Weekday* for variable day of week, D_w and 24 values *A, B, ..., X* representing 24 hours in a day, for variable hour of day, H_d. The parameters mentioned in the previous point are also converted from continuous to discrete attributes(more in Section 3.3), to make a uniform dataset with discrete attributes. The list of all the available parameters is shown in Table 3. They are categorized into 3 domains: DHS operational parameters, weather forecast parameters and behavioural parameters.

Table 3: Parameters considered for forecasting heat load.

DHS Operational Parameters	Weather Forecast	Behavioral Parameters
Supply temperature(T_s) Return temperature(T_r) Flow rate(m) Difference of supply and return temperature(T_{delta})	Outdoor Temperature (T_{out})	Hour of Day(H_d) Day of Week(D_w)

The objective of this paper is to predict the heat load forecast using the available influencing parameters. We are not interested in studying the dependencies between the influencing parameters. However, it is also possible to study the relationships among different parameters by modelling a complex Bayesian network. But to predict the heat load forecast it is enough to study the influence of each parameter on the heat load forecast individually. Therefore, we decide to model our Bayesian network as Naive Bayes network as it learns efficiently and provide accurate forecasts. Though it is a simple model, it has proven to work well by achieving similar performances with state-of-the-art methods [4]. We also validate this assertion in this paper in section 4.

3.3 Parameter Discretization

We discretized heat load and other influencing parameters for learning our Bayesian network as shown in Figure 4. As each building has its own unique trend of heat load variation (see tables 2 and 3), the discretization was carried out separately for each building. The discretization process was automated to reduce the manual overhead of making states for each parameter by the expert. This was carried out by using unsupervised discretization techniques: Equal width discretization(EWD) [2] and k-means clustering [7].

Equal Width Discretization: The continuous values corresponding to a numeric attribute are first sorted from minimum (v_{min}) to maximum (v_{max}) values and then divided into k intervals of equal width w, where k is a user supplied parameter.

$$w = (v_{max} - v_{min})/k \quad (3)$$

K-means Clustering: divides the dataset into k clusters, where k is fixed in advance. The underlying algorithm is explained in detail in [7]. k-means is applied to all DHS operational parameters and weather forecast parameters individually for discretization. In terms of discretization, each cluster represents a discrete state with a well-defined range. For example, a *cluster0* corresponds to a discretized state,

State A with a range of *[a,b]* where a and b are the continuous attributes belonging to the dataset, such that $a<b$. The next cluster, *cluster1* corresponds to State B with a range *[c,d]* where $c>b$ and $c<d$.

3.4 BayesForSG: A Bayesian Model for Forecasting Thermal Load in Smart Grids

Figure 3 shows the proposed model for forecasting the heat load by utilizing the district heating data collected from Buildings A, B and C. The DHS operational parameters, heat load consumption and outdoor temperature are collected at a per minute interval by the sensors deployed at the substations in different buildings. This data is collected together and converted to a resolution of per hour, to take into account the fact that the delay in the control loop of the Skellefteå Kraft is between 4 to 6 hours. Then k-means and EWD are used to discretize the heat load and each continuous parameter into discrete states with well-defined ranges. The discretized parameters serve as an input to the Naive Bayes classifier which classifies the forecasted heat load to a particular state. The output value of the heat load forecast serves as input to the production grid of Skellefteå Kraft.

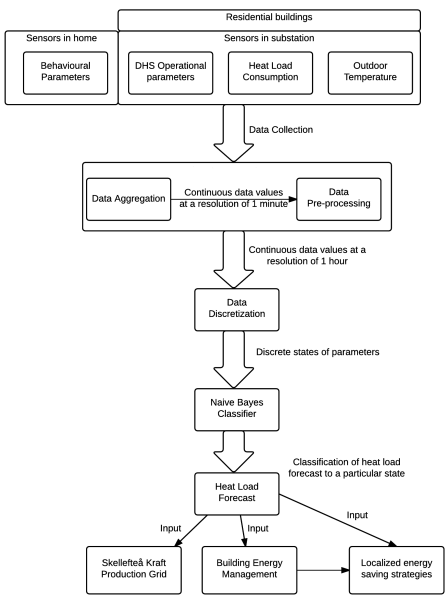

Figure 3: The proposed model for forecasting heat load.

For modelling the Bayesian network we use the notation as shown in Table 3 for the random variables. The current time is represented at t and the load forecast has to be computed for $t+h$ hours into the future. Here h is the forecast horizon. The current outdoor temperature: $T_{out(t)}$, outdoor temperature forecast at $t+h$ hours: $T_{out(t+h)}$, current heat load: $HL_{(t)}$, heat load forecast at $t+h$ hours: $HL_{(t+h)}$. A generic Naive Bayes network representing the heat load forecast using all the available parameters is shown in Figure 4. Consider a set $A = \{A_i, ..., A_n\}$ which contains the parameters A_i to A_n that influence the heat load forecast $HL_{(t+h)}$, while being conditionally independent from each other. This relationship is depicted in the Naive Bayes model as shown below:

$$P(HL_{(t+h)}|A_i) = \arg\max P(HL_{(t+h)}) * \prod_{i=1}^{n} P(A_i|HL_{(t+h)}) \quad (4)$$

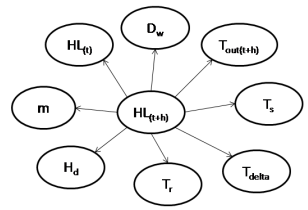

Figure 4: Naive Bayes network for heat load forecast using all available parameters.

We model the Naive Bayes network for heat load forecast by considering four cases as discussed below, by carrying out trace driven analysis in Weka [3]. The behavioural parameters hour of day, H_d and day of week, D_w are added in all the cases to consider the effect of user behaviour. These include: I.) influence of DHS operational parameters on the heat load forecast; II.) influence of outdoor temperature forecast on the heat load forecast; III.) influence of DHS operational parameters and outdoor temperature forecast on the heat load forecast; and IV.) influence of current heat load consumption and outdoor temperature forecast on the heat load forecast. These cases are discussed in the next section in detail along with the results.

4. RESULTS ANALYSIS

In this section, we study the results of the heat load forecast using our proposed model in three residential buildings over a period of winter and spring seasons. The model was evaluated using 10 folds cross validation over the horizons of 1, 2, 3, 6 and 24 hours for all four cases mentioned in the previous section. For evaluating the proposed model, it is necessary to compute classification accuracy. The accuracy highlights the ability of the model to correctly predict states and also to differentiate among the states. Since we are computing the accuracy of the heat load forecast, we will refer it as forecast accuracy. It is defined by the equation below.

$$Forecast\ accuracy = \frac{Number\ of\ correct\ classifications}{Total\ number\ of\ instances} \quad (5)$$

We now present the results of heat load forecast individually for each case.

4.1 Results evaluation

Case I: Influence of DHS operational parameters: The Naive Bayes network for this case is shown in Figure 4 by considering DHS operational and user behavioural parameters: T_s, T_r, m, T_{delta}, H_d and D_w and their influence on the heat load forecast $HL_{(t+h)}$. This Bayesian network classifies the state of $HL_{(t+h)}$ by assuming conditional independence among the influencing parameters. The state of

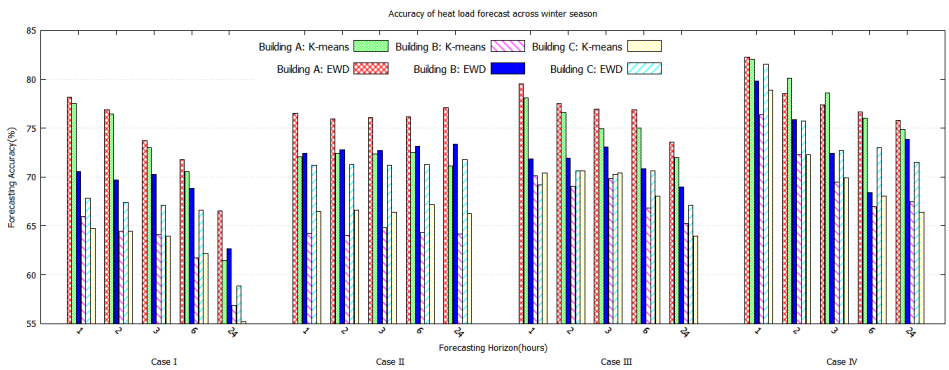

Figure 5: Accuracy of heat load forecast across winter season.

$HL_{(t+h)}$ is determined by choosing the state of heat load with the highest probability. The accuracies of heat load forecast across both winter and spring seasons is presented in Figures 5-6. The results of forecasting accuracy for Case I indicate that the accuracy of the heat load forecast mostly decreases with the increase in forecasting horizon for a particular building. This can be explained by Eq. 2. The magnitude of the heat load consumption at a substation depends on the current values of DHS operational parameters which are used for forecasting heat load using our trained Bayesian network. With the increase in the forecasting horizon, the dependency of the heat load forecast decreases on the current values of DHS operational parameters, which results in the decrease in accuracy.

Case II: Influence of outdoor temperature forecast: In this case, the state of the heat load forecast is predicted by considering the influence of outdoor temperature forecast and user behavioural parameters. The Naive Bayes network is represented by considering $T_{out(t+h)}$, H_d, D_w and $HL_{(t+h)}$ from Figure 4. The results of Case II indicate that in both winter and spring seasons, the accuracy of the heat load forecast across various horizons almost remains constant for all buildings. This is possibly due to the fact that the heat load forecast at different horizons has similar dependency on the outdoor temperature forecast.

Case III: Influence of DHS operational parameters and outdoor temperature forecast: In this case, we consider the influence of DHS operational parameters and outdoor temperature forecast on the heat load forecast. The Naive Bayes network is depicted by considering T_s, T_r, m, T_{delta}, H_d, D_w, $T_{out(t+h)}$ and $HL_{(t+h)}$ from Figure 4. During the spring season, the accuracy of the heat load forecast decreases with the increase in forecasting horizon in majority of the cases for a particular building. This decrease is explained by Eq.2 due to the presence of DHS operational parameters as discussed in Case I. In winter season, there was no clear trend in the accuracy of heat load forecast across various horizons due to the stochastic nature of heat load variation.

Case IV: Influence of current heat load consumption and outdoor temperature forecast: The Naive Bayes network considering the influence of current heat load consumption and outdoor temperature forecast on the heat load forecast is represented by choosing the parameters $HL_{(t)}$, $T_{out(t+h)}$, H_d, D_w and $HL_{(t+h)}$ from Figure 4. The results of Case IV indicate that during winter season, the accuracy for $HL_{(t+1)}$ is higher than the accuracy of other horizons for all buildings. This indicates that the current load consumption has a strong influence on the heat load forecast for the next hour. However, with increasing forecasting horizon, the dependency of heat load forecast on the current heat load decreases which results in a decrease in forecasting accuracy. This is because the the heat load variation increases with time. For instance, there will be a small difference in heat load consumption in consecutive hours. However, there will be a much higher difference in heat load consumption in 6 hours. Thus, current heat load consumption influences the heat load forecast with the varying forecasting horizon. In the spring season, in most of the cases, the forecasting accuracy increases from 6 hour to 24 hour horizon. This is because of user activity pertaining to a daily routine, which influences the heat load forecast.

4.2 Discussion

The results from Figures 7-8 indicate that Case IV achieves higher average accuracy than other cases for both seasons with both discretization techniques. This implies that current heat load consumption and outdoor temperature forecast are the two parameters with most influence on the heat load forecast. The results from Figures 5-6 indicate that EWD performs better than k-means in majority of the cases. We observe that k-means performs well and at par with EWD in buildings (Building A) with less variation of heat load consumption. However, EWD performs better than k-means in buildings (Buildings B and C) with higher heat load variation. These two discretization techniques need to be tested on more buildings in order to compare their impact on the forecasting accuracy. Building A achieves a higher forecast accuracy over all horizons across both seasons in all four cases. This is due to the fact that Building A has much less variation in heat load consumption in both seasons as compared to buildings B and C as shown in Tables 2 and 3. Thus, the proposed model forecasts heat load consumption with a higher accuracy in case of less heat load variation.

The proposed model achieves best case average accuracies of 81.23% and 76.74% for a forecast horizon of 1 hour($HL_{(t+1)}$) in the three buildings for winter and spring seasons respec-

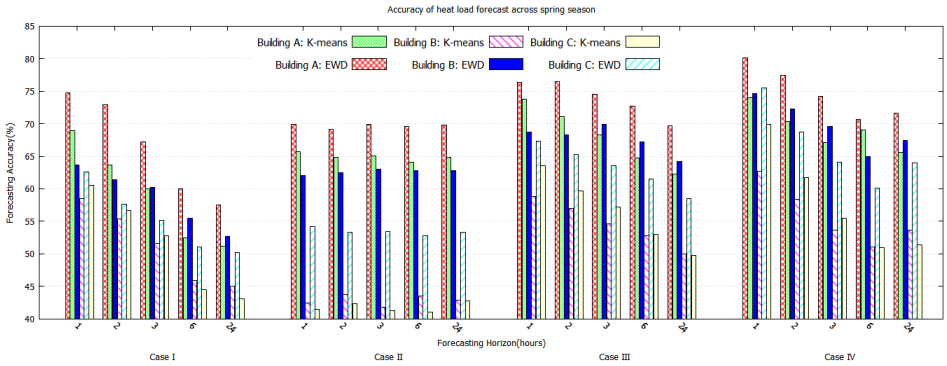

Figure 6: Accuracy of heat load forecast across spring season.

tively in Case IV using EWD. The model also achieves an average best case accuracy of 77.97% using EWD (for all three buildings together) across both seasons for the forecast horizon of 1 hour($HL_{(t+1)}$) by utilizing only 10% of the training data. Figure 9 shows the accuracy of $HL_{(t+1)}$ forecast for different percentages of training data for Buildings A, B and C over the period of both seasons. These results indicate that the proposed model satisfies the first requirement for using less training data to build efficient forecast models. We observe that the resulting accuracy is almost similar for different percentages of training data. BayesForSG achieves high forecast accuracies with just 10-20% training data. It doesn't require large amount of training data to achieve higher accuracies.

4.3 Comparison of BayesForSG with state of the art methods

Figure 10 shows the comparison of BayesForSG with state of the art methods for heat load forecast including multilayer perceptron(MLP), support vector machine(SVM) and Random forest. These methods are evaluated using different percentages (10%, 30%, 50% and 70%) of training data to compute $HL_{(t+1)}$ using EWD in Case IV(best case scenario). We compute their average accuracies for all three buildings across both seasons. These results indicate that BayesForSG achieves higher average accuracy than other methods across various percentages of training data. Also, by utilizing only 10% training data BayesForSG outperforms other methods by a significant margin. With the increase in training data, the performance of other methods improve while that of BayesForSG almost remains constant. The accuracy of BayesForSG is in the range of 77.558% to 78.065% across various percentages of training data. This shows that BayesForSG doesn't require large amount of training data to achieve high accuracies. It achieves high accuracy by just utilizing 10% training data. This makes BayesForSG a more efficient model than other methods for thermal load forecast. Thus, BayesForSG satisfied the requirement of using less training data by providing more accurate results than state of the art methods for heat load forecast. It was also successful to forecast the heat load with a good accuracy in three different buildings with different heat load variations. This shows that BayesForSG can be applied for applications of heat load forecasting in different buildings.

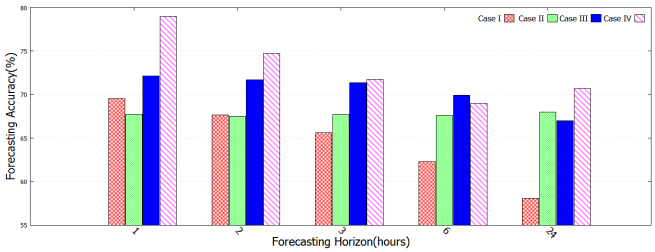

Figure 7: Average forecasting accuracy(%) for all four cases across both seasons for Bldgs. A, B and C using EWD discretization and Naive Bayes Network.

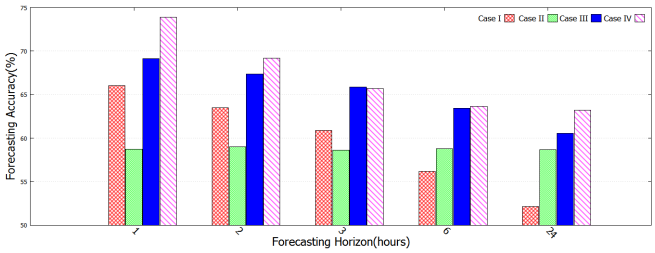

Figure 8: Avg. forecast accuracy(%) for all four cases across both seasons for Bldgs. A, B and C using k-means clustering and Naive Bayes network.

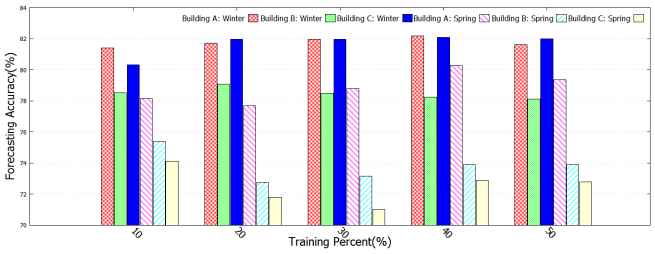

Figure 9: Avg. forecast accuracy for $HL_{(t+1)}$ with different percentages of training data using EWD and Naive Bayes network in Case IV.

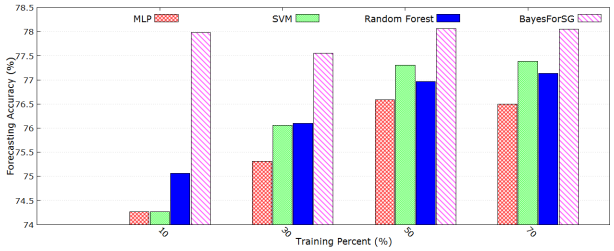

Figure 10: Avg. forecast accuracy of MLP, SVM, Random Forest and BayesForSG for $HL_{(t+1)}$ for all three buildings across both seasons using EWD in Case IV for different percentages of training data.

5. CONCLUSION AND FUTURE WORK

This paper presented an efficient model BayesForSG, for forecasting heat load consumption in a smart grid environment. We modelled a Naive Bayes network to study the influence of various parameters using four cases. The forecast model was built by utilizing the realistic district heating data over a period of 4 months across winter and spring seasons from three residential buildings in Skellefteå, Sweden. Heat load forecasting was performed for horizons of 1, 2, 3, 6 and 24 hours to consider the effect of the district heating control loop and daily heat load consumption pattern. Our results indicate that the current heat load consumption and outdoor temperature forecast are the two most important parameters influencing the heat load forecast. In this case, our model achieves average accuracies of 81.23% and 76.74% for a forecast horizon of 1 hour in three buildings for winter and spring seasons, respectively. Further, by utilizing only 10% of training data, BayesForSG was able to achieve an average accuracy of 77.97% for the three buildings across both seasons with forecast horizon of 1 hour. BayesForSG also outperformed state of the art methods for thermal load forecast by achieving higher average accuracies in all three buildings across both seasons.

The future research may benefit by incorporating the influence of additional weather parameters(wind speed, wind direction, humidity, solar radiation etc.), building characteristics and occupant behaviour on heat load forecast.

6. ACKNOWLEDGMENTS

The authors would like to thank Erasmus Mundus PERCCOM program(Pervasive Computing and Communications for Sustainable Development) and Erasmus+ program of European Commission for providing funding for this research. The research is a part of the FP7 Project, OPtimising Hybrid Energy grids for smart cities(OrPHEuS), grant agreement 608930, funded by the European Commission.

7. REFERENCES

[1] T. Catalina, V. Iordache, and B. Caracaleanu. Multiple regression model for fast prediction of the heating energy demand. *Energy and Buildings*, 57:302–312, 2013.

[2] J. Dougherty, R. Kohavi, M. Sahami, et al. Supervised and unsupervised discretization of continuous features. In *Machine learning: proceedings of the twelfth international conference*, volume 12, pages 194–202, 1995.

[3] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[4] J. Huang, J. Lu, and C. X. Ling. Comparing naive bayes, decision trees, and svm with auc and accuracy. In *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*, pages 553–556. IEEE, 2003.

[5] S. Idowu, S. Saguna, C. Ahlund, and O. Schelén. Forecasting heat load for smart district heating systems: A machine learning approach. In *Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on*, pages 554–559. IEEE, 2014.

[6] B. Lapillonne, C. Sebi, K. Pollier, and N. Mairet. Energy efficiency trends in buildings in the eu. *Lessons from the ODYSSEE/MURE project. ADEME, Supported by Intelligent Energy Europe. Online at: http://www. odysseeindicators. org/publications/PDF/Buildings-brochure-2012. pdf*, 2012.

[7] J. MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA., 1967.

[8] E. Parliament. Directive 2010/31/eu of the european parliament and of the council of 19 may 2010 on the energy performance of buildings, 2010.

[9] M. Protić, S. Shamshirband, M. H. Anisi, D. Petković, D. Mitić, M. Raos, M. Arif, and K. A. Alam. Appraisal of soft computing methods for short term consumers' heat load prediction in district heating systems. *Energy*, 82:697–704, 2015.

[10] R. Ranjan. Modeling and simulation in performance optimization of big data processing frameworks. *Cloud Computing, IEEE*, 1(4):14–19, 2014.

[11] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Pearson Education, Inc. and Dorling Kindersley Publishing, Inc., second edition, 2009.

[12] F. Sandin, J. Gustafsson, and J. Delsing. Fault detection with hourly district energy data. *Stockholm: Swedish district heating association*, 2013.

[13] M. Vlachopoulou, G. Chin, J. C. Fuller, S. Lu, and K. Kalsi. Model for aggregated water heater load using dynamic bayesian networks. Technical report, Pacific Northwest National Laboratory (PNNL), Richland, WA (US), 2012.

[14] M. Wang and Q. Tian. Application of wavelet neural network on thermal load forecasting. *International Journal of Wireless and Mobile Computing*, 6(6):608–614, 2013.

[15] K. Wojdyga. Predicting heat demand for a district heating systems. *International Journal of Energy and Power Engineering*, 3(5):237–244, 10 2014.