OPTIMIZED PARAMETER SELECTION FOR ASSESSING BUILDING ENERGY EFFICIENCY

Elena MOCANU *, Phuong H. NGUYEN*, Madeleine GIBESCU* and Wil L. KLING*

*Eindhoven University of Technology, Department of Electrical Engineering, 5600 MB, Eindhoven, The Netherlands

E-mail: e.mocanu@tue.nl, p.nguyen.hong@tue.nl, m.gibescu@tue.nl, w.l.kling@tue.nl

Abstract. This study is analyzing the influence of various physical parameters on the heating and cooling load requirements of a building. In order to determining the most relevant combination of parameters for assessing energy efficiency we have created a framework which combines a stochastic classification method, namely Gaussian Mixture Model, with a combinatorial optimization procedure. The framework was evaluated using a simulated benchmark database consisting of 768 buildings.

Keywords: Building Parameters, Energy efficiency, Gaussian Mixture Model, Optimization.

INTRODUCTION

This study looks into the problem of assessing the building energy efficiency. More exactly the heating load (HL) and cooling load (CL) are studied as a function of a number of physical building parameters. The optimal number of the parameters used to predict the energy consumption required for HL and CL, calculated on a yearly basis, were found using an optimization procedure.

PROBLEM DEFINITION

In recent years, a large number of approaches have been proposed to model trends in energy consumption for future buildings systems [1-5]. In general these methods are Supervised Learning (SL) methods. In SL, one set of observations, called inputs, are assumed to be the cause of a different set of observations, called outputs. SL is the machine learning procedure which tries to discover a function from labeled training data. More formally, if we define an input space X and an output space Y, the question of learning is reduced to the question of estimating a functional relationship of the form C: X →Y that is a relationship between input and output. Such a mapping C is called a classifier. The flow diagram for this procedure can be seen in Figure 1.

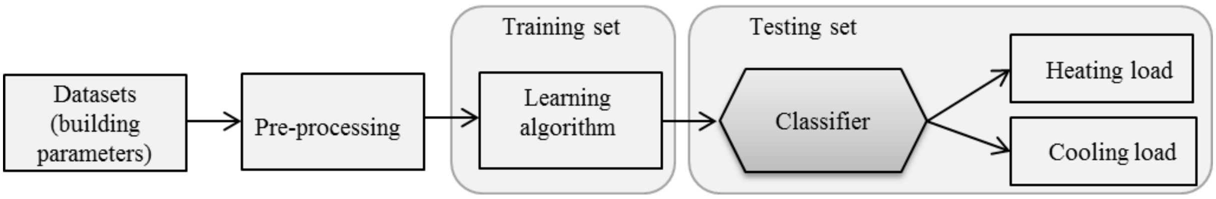

Figure 1. Supervised learning flow diagram.

In our specific case, the X space is given by the building parameters and the Y space is given by the HL and CL. Further the mathematical details of the classifier used are presented below.

GAUSSIAN MIXTURE MODEL

A Gaussian Mixture Model (GMM) is defined as a convex linear combination of several probability density functions represented as a weighted sum of M Gaussian component densities given by $P(x \mid \lambda) = \sum_{i=1}^{M} w_i g(x \mid \mu_i \Sigma_i)$ where x is a D -dimensional continuous-valued data vector. w_i , $\forall i = 1 \dots M$ are the mixture weights, and $g(x \mid \mu_i \Sigma_i)$ $\forall i = 1 \dots M$ are the component Gaussian densities. Each component density is a D-variant Gaussian function with mean vector μ_i and covariance matrix Σ_i . The mixture weights satisfy the constraint that $\sum_{i=1}^{M} w_i = 1$.The complete Gaussian mixture model is parameterized by the mean vectors, covariance matrices and mixture weights from all component densities. These parameters are collectively represented by the notation $\lambda = (w_i, \mu_i, \Sigma_i)$, $\forall i = 1 \ldots M$. Additionally, parameters can be shared among the Gaussian components determining a common covariance matrix for all features. The linear combination of diagonal covariance basis Gaussians is capable of modeling the correlations between feature vector elements. One of the powerful attributes of the GMM is its ability to produce smooth approximations to arbitrarily shaped densities. The classical uni-modal Gaussian model represents feature distributions by a position and an elliptic shape. A vector quantizer (VQ) or nearest neighbor model represents a distribution by a discrete set of characteristic templates [5]. A GMM acts as a hybrid between these two models by using a discrete set of Gaussian functions, each with their own mean and covariance matrix, to provide a better modeling capability.

EXPERIMENTS AND RESULTS

The benchmark dataset [6] used in the experiments comprises 768 samples. Each sample has 8 features to predict the real valued responses (i.e. HL and CL). These responses can also be seen as a multi-class classification problem if their values are rounded to the nearest integer. Fig. 2 shows the distribution of the classes. Overall, the samples represent 12 different building shapes, with the same volume, simulated in Ecotect. They differ with respect to the following parameters: Relative Compactness(X1), Surface Area (X2), Wall Area(X3), Roof Area(X4), Overall Height (X5), Orientation (X6), Glazing Area (X7), Glazing Area Distribution(X8). The aim is to find the optimal combination of the parameters in order to achieve the best classification accuracy of HL and CL. In Table 1 are presented the combination of the parameters which perform with the highest accuracy. The GMM accuracy using all parameters is 71.61% for HL and 73.82% respectively for CL.

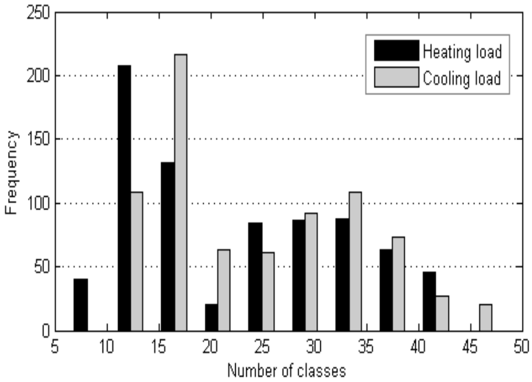

Figure 2. Heating load (HL) and cooling load (CL) rounded to the nearest integer

Table 1. Classification accuracies for GMM using the optimal parameters

X_i(HL)	Accuracy	X_i(CL)	Accuracy
[2,5,7]	73.6%	[1,2,3,7,8]	75.1%
[2,5,8]	73.6%	[1,3,4,7,8]	75.1%
[2,5,6,7]	73.6%	[1,2,3,6,7,8]	75.1%
[2,5,6,8]	73.6%	[1,3,4,6,7,8]	75.1%
[2,5,7,8]	73.6%	[2,3,7]	74.9%
[2,5,6,7,8]	73.1%	[2,3,8]	74.9%
[2,5]	72.6%	[3,4,7]	74.9%
[2,3,7,8]	72.6%	[3,4,8]	74.9%
[1,3,5,7,8]	72.6%	[2,3,6,7]	74.8%
[2,3,5,7,8]	72.6%	[2,3,6,8]	74.8%

CONCLUSIONS

The exhaustive search was performed using the GMM output for each combination of the parameters. The results identify the most influential variable or group of variables highlighting the optimal choice of parameters for modeling heating and cooling loads in a building. The accuracy obtained using only three parameters, namely “Surface Area”, “Overall Height” and “Glazing Area”, show an improvement over the accuracy using all data for HL. Also, a combination of five physical parameters of the building showed a higher accuracy in predicting CL than using the entire dataset. These results, somewhat counter-intuitive, are thought to improve the computation time for optimization models aiming to achieve more energy-efficient buildings.

REFERENCES

[1] N. Djuric and V. Novakovic: "Identifying important variables of energy use in low energy office building by using multivariate analysis ", Energy and Buildings, Vol. 45, pp. 91-98, 2012

[2] A. Foucquier, S. Robert, F. Suard, L. Stphan, and A. Jay, “State of the art in building modelling and energy performances prediction: A review,” Renewable and Sustainable Energy Reviews, vol. 23, no. 0, pp. 272 – 288, 2013

[3] H. xiang Zhao and F. Magouls, “A review on the prediction of building energy consumption,” Renewable and Sustainable Energy Reviews, vol. 16, no. 6, pp. 3586 – 3592, 2012.

[4] M. Aydinalp-Koksal and V. I. Ugursal, “Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector,” Applied Energy, vol. 85, no. 4, pp. 271 – 296, 2008.

[5] R.Gray: Vector Quantization, IEEE ASSP Magazine, pg.4–29, 1984

[6] A. Tsanas, A. Xifara: "Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools", Energy and Buildings, Vol. 49, pp. 560-567, 2012