Simulation of Air Conditioning System for Thermal Comfort Using Machine Learning Techniques

Dattatray L. Bankar[*,1], Sandesh K. Rasal[2], Amar P. Pandhare[3]

[1,2]*Assistant Professor, Sinhgad College of Engineering, Pune-41*
[3]*Professor, Sinhgad College of Engineering, Pune-41*

Abstract: *This study is about the system which establishes a model for environmental factors such as temperature, humidity, and air velocity of thermal comfort indicators by implementing different Machine Learning technique. In this work the principal goal is analyzing air flow, distribution of temperature and pressure in the Living room. In order to achieve the goal, the domain is considered which is divided in to discretized manner which will be further used to record the thermal sensation and airflow. For the same simulated environment, the experiment will be conducted, which resulting in to forming the dataset with respect time, relative humidity, draft, temperature, air flow. By considering the recorded dataset, machine learning algorithm like SVM (Support Vector Machine), Liner Regression, XGBoost are used to create a model in Python. Finally, these algorithms evaluated with various performance metrics on basis of Mean Squared Error (MSE) method. The algorithm which has least MSE is best suited algorithm.*

Keywords: Air Conditioning System, Thermal Comfort, Machine Learning Algorithms, Simulation, Boosting techniques.

1. Introduction

Air conditioning is used in numerous sectors worldwide to provide the comfort as well as appropriate environment to occupants. The air conditioning is the branch of engineering science which deals with the study of conditioning of air i.e., supplying and maintaining desirable internal atmospheric conditions for human comfort, irrespective of external conditions, which elaborate as “the condition of mind, which express satisfaction with thermal environment”. when people are not satisfied with surrounding thermal environment then it can affect to occupants’ productivity and work efficiency through several ways like it will cause attention distraction and reduce the ability to concentrate properly and sometimes it also affects to human health. So, maintaining the constant thermal comfort is important consideration while designing the air conditioning system [1].

There are several factors which affects the thermal comfort. All factors are independent with each other but together they affect the thermal comfort at great extent, which mainly classified in to two domains. The first one is Environmental factors like humidity, temperature, motion of air and purity of air and consecutively personal factors like type of clothing, age and sex, duration of stay, density of occupants, number of heating and electric appliance in room [2].

The traditional control systems are not able to meet the requirement of human comfort so there is demand for improved technologies like AI and machine learning for more

accuracy and efficiency and optimization of the parameters. Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being actual conducting the program. Machine learning algorithms use historical data as input to predict new output values. Machine learning algorithms can be classified into three main learning types: Supervised learning: In supervised learning, along with independent attribute the dependent attribute is also underline irrespective of the regression and classification problem. Where as in case of the Unsupervised learning the dataset having only independent attribute which certainly known as not labeled data [5,6]. Depending on the type of labels, the learning approach differs. the aspect with which the data is distributed decides the type of algorithm to be implemented. The statistical technique which measures central tendency of the data, which majorly helps to analyze the data and find out the correlation between them. Correlation like Pearson and Spars man is contributing for the same. Along with that it also helps in reducing the curse of dimensionality [5,7]. With different parameters of building environment by keeping in mind the thermal sensation of the occupants at the same time optimization of the energy consuming equipment's. Support vector machine (SVM) with Different kernels, Decision tree redresser, gradient boosting with its Adaptive and extreme nature, implemented on the mentioned independent attribute. Which simulate the environment accordingly. [5,6,7,8,9]

The Performance metrics which involve in the domain, Either Supervised or unsupervised are calculated based on the Actual value and Predicted with the help of algorithm. the standard practice, underline that, split the dataset in training and testing set. The model is train on the split training dataset, the performance of the model is evaluated on spilt testing dataset. Mean absolute error (MAE), mean squared error (MSE), log mean squared error (LMSE), Comparing the accuracy metrics with different algorithm, one can conclude which algorithm is best for which dataset [3,4,5,6,7,8]. It is not limited to the above approach, the problems of the prediction model are addressed by artificial neural network (ANN) which consist of the input layer, hidden layer, output layer, the transfer function like sigmoid, tanh, SoftMax, rely, which provide the non-linearity in the data. The optimizer function like gradient descent, stochastic gradient descent, ad grad, adam, help to increases the accuracy of the mode [8,9.10,11]. By selecting the proper features from the data is vital part in building the model, the statistical approach is suitable for the same. The metrics like covariance and correlation calculated and plotted with the help of seaborn library, which identified the considerable features impact the prediction model [10,11,12]. By considering the split air conditioning system which helps to generate the cooling environment for residential purpose. The proposed study consists of the typical residential room as mentioned. the objective of the context is to implement the supervised learning model by taking in to account the space, the density of the occupants, majorly the velocity with which air flows and humidity condition for the focused environment. Typically, study implements the supervised learning algorithm, Support Vector Machine (SVM), Random Forest (RF), Boosting algorithm.

2. Methodology

The Major focus in the Study is to develop the prediction-based model which simulate the cooling environment, in order to identify the draft and thermal sensation and comfort for the human being. In doing so the process is adopted in five modules through which the basic framework was created, which certainly helps to enhance the control reliability of the involved parameters.

Step 1: Problem framing. The first step is to have the clarity in the Aspect of determination of the model and its problem specification, boundary condition in terms of the environment, assumptions for the same, the aligned previous data with considering the parametric metrics.

Step 2: Data preparation. The second step is to collect the data, for the current study involved, data acquisition technique is used with measuring the temperature and humidity and velocity sensors placed at respective zone. Due to sensor failure some data may be missing or some raw get introduced in the zone, with the help of the statistical technique (mean, median and mode) where deciding criterion is its histogram of that particular attribute.
Step 3: Learning: Based on the prepared dataset, different learning model is implemented. The training on the data with different regularization technique will help to build the more robust model, suitable cross validation technique which average determine the training accuracy and avoid the randomness in the accuracy metrics.
Step 4: Prediction: The test data which introduced from the train test split, was used to check the performance of the model by different loss function which actually designate the accuracy of the model and also allow for modification of the model if it is not up to the mark.
Step 5: Validation: By improving the problem of the performance metrics, may some time the issue of bias and variance is arise which make model to behave unfairly since bias is more, by keeping in mind the bias and variance trade off and hyper tuning parameter, Ones the model become robust. Then it comes for the validation, by providing the unknown attribute to them, the data is predicted and match with the data acquisition module.

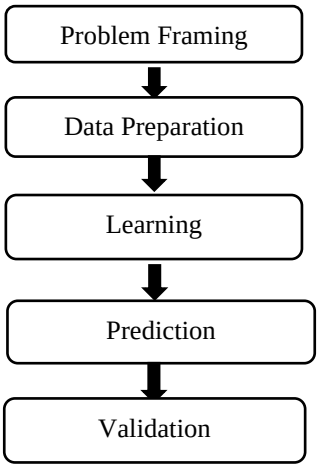

Figure 1. Flow Chart

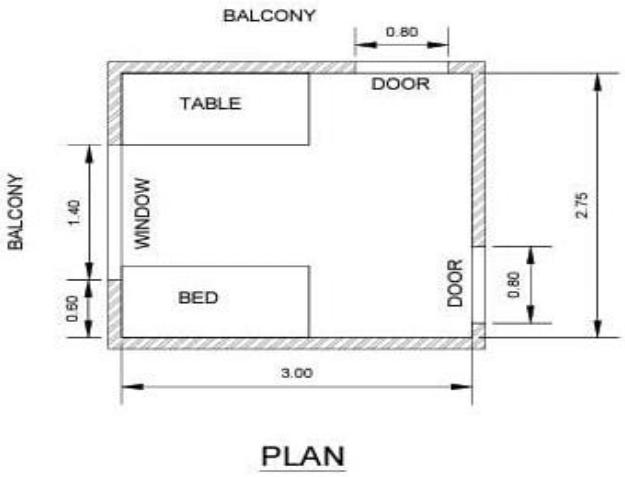

Figure 2. Floor Layout

Table 1. Room Dimensions

Parameters	Room Dimensions (meter)
Height	3.2
Width	2.7
Length	3

Table 2. Load & its Parameters

Load	Parameter
Tube-light	20 Watt
Roofing	Concrete (k=20.82 W/m k)
Flooring	Tile (k=20.82 W/m k)
Concrete Wall	4 sides
Load	2-4 person
Number of air changes per hr.	Two walls exposed=1.5

3. Machine Learning

Observations, experience, analysis, experimental sets, are computationally aligned in the structured ways with different statistical technique which intern predicts output of the model with given attribute [3], the data driven behaving of the statistical measuring parameter clubbed with respect label data and model building is performed. Knowing the output of the given featured it is further classified in the supervised and unsupervised sense. The algorithm which used in the current study is only discussed below.

3.1 Support vector Machine

Mapping of the original dataset with the features space F by using the statistical techniques gives rise to the birth of SVM. For Implementation of SVM in classification and regression problem with Linear as well as nonlinear separable data by using different kernels, Hyperplane plays important role, which actually led the map of SVM. along with Hyperplane, two vectors on both sides, helps to create the Maximum Marginal distance, on which the performance depends. Mainly if the data is Linear Separable, the Linear

kernel is Suitable built up the model, but if dataset is not accompanied with Linearity the Kernals (Radial Basis, polynomial) are introduced to build the most robust model for the same. For the less no of training sample in the domain the SVM gives the better accuracy which makes id suitable to use in any generalized condition.

$$f(x) = \langle \omega, x \rangle + b \; with \; \omega \in \mathbf{X}, b \in \mathbb{R}$$

$$|\xi|_\varepsilon := \begin{cases} 0 & if \; |\xi| \leq \varepsilon, \\ |\xi| - \varepsilon & otherwise. \end{cases}$$

$$k(x, x') := \int_\chi S(x, z) s(x', z) \, dz$$

3.2 Decision tree

The Randomness of the features is basis of working on the DT. The dataset feature extraction and creation of the root node based on some decision is the basic working environment of DT. The decision for the section of feature for corresponding node is based on the team GINI, which again based on Entropy of the Feature. Entropy defines the randomness in the attribute. The term variance is implemented to calculate the error of each attribute that ill deciding factor for the decision tree regressor.

$$\mathrm{IG}(D_p, f) = I(D_p) - \left(\frac{N_{left}}{N_p} I(D_{left}) + \frac{N_{right}}{N_P} I(D_{right}) \right)$$

3.3 Random Forest

Based on the Ensemble Bagging technique which selects the random sample and that too with replacement. Each random dataset is fed to separate decision tree and club of all such tree refer the mentioned algorithm. Bagging is self is parallel learning provides the voting algorithm at the end, for regression problem statement the average is taken as output the given attribute prediction. The terms involve to deciding the Root and node features will be reflection in random forest also since it combining the DT only. The CART algorithm takes the GINI as parametric term since its value varies from 0 to 0.5.

3.4 Boosting Technique

The weak learner output is basis of all boosting technique. Stumps are created and arranged sequentially, Incorrect information is Passed through sequence of Stump's which leads to minimization of error at the end, it is based on the strategy of learning through mistakes. Boosting techniques are classified in to three categories

1. Gradient Boosting
2. Adaboost
3. Xtreme Gradient boosting

Gradient boosting algorithm is based on the Pseudo code, where base model is created on the basis of loss function, on the basis of the base model The residual is calculated, for the given number of the Sequential weak learner, the residual and the attribute are fitted over the weak learner, the Optimizer which involves in pseudo code makes the Gradient Boosting model strong.

Input: training set $\{(x_i, y_i)\}_{i=1}^n$, a differentiable loss function

In case of the Adaboost, Equal weights are assigned for each of the Attribute, Th incorrect sample Weights are added based on which the parameter like performance of Stums is Calculated for each time the Weights are updated, For Incorrect Attribute the Weight updated ion higher side and reverse for Correct set. Which every time updated the Weight

unless and until the error is reached to minimum through sequential learning, the normalization of the weight is key parameter in Adaboost.

Given: ((x_1, y_1) , ……. (x_m, y_m) where $x_i \in X, y_i \in Y = \{-1, +1\}$

Initialize $D_i(i)$ =1/m

For t = 1, …., T:

- Train weak learner using distribution D_t.
- Get weak hypothesis $h_t : X \to \{-1, +1\}$ with error

$$\epsilon_t = Pr_{i \sim D_i}[h_l(x_i) \neq y_i].$$

- Choose $\propto_t = \frac{1}{2} In \left(\frac{1-\epsilon_t}{\epsilon_t}\right)$.
- Update:

$$D_{t+1}(i) = \frac{D_{t(i)}}{Z_i} \times \begin{cases} e^{-\alpha_t} & if\ h_t(x_i) = y_i \\ e^{\alpha_t} & if\ h_t(x_t) \neq y_i \end{cases}$$

$$= \frac{D_t(i) \exp(-\alpha_t y_t h_t(x_i))}{Z_t}$$

Where Z_t is normalization factor (chosen so that D_{t+1} will be distribution).

Output the final hypothesis:

$$H(x) = sign(\sum_{t=1}^{T} \alpha_t h_t(x))$$

4. Results & Discussion

The performance metrics implemented fir the comparing the different set of algorithms is mean squared error (MSE) which takes care of both actual and predicted value. The parameter consider for prediction of temperature and Humidity is the location i.e., x, y, z. and velocity.

1. Initially the Linear Regression is implemented for given training dataset. It is observed that after simulation of all 4 algorithms. the MSE is not much deviated from its Value.
2. From values of MSE for all Algorithm, the value is minimum for Linear Regression.
3. The Boosting Algorithm almost fluctuating in very limited manner.

Table 2. Load & its Parameters

Parameter	Linear Regression	SVM	XGBoost
MSE(Temp)	14.5	14.44	14.49
MSE(Velocity)	0.08	0.088	0.078
MSE(Humidity)	4.280	4.455	4.451
Mean Error	6.28	6.32	6.33

5. Conclusion

Consideration to the different Machine learning Technique, looking towards the loss function as MSE, the Linear Regression take lowest value of the error rate leads to higher

accuracy value among rest algorithm for prediction of temperature, humidity and velocity. Support Vector Machine bound to be less accurate for the prediction model. We can use this model to predict the location of Air Conditioning to achieve thermal comfort and ultimately increase the efficiency of air conditioning system. This model will be useful for designing air conditioners for home, theatres, malls and auditoriums.

REFERENCES:

[1] *Peng, Yuzhen & Rysanek, Adam & Nagy, Zoltán & Schlüter, Arno, 'Using machine learning techniques for occupancy-prediction-based cooling control in office buildings,' Applied Energy, Elsevier, vol. 211(C), (2018) pp. 1343-1358.*

[2] *Yafei Zhao, Paolo Vincenzo Genovese, Zhixing Li, "Intelligent Thermal Comfort Controlling System for Buildings Based on IoT and AI," Future Internet, MDPI, vol. 12(2), (February 2020) pp. 1-18.*

[3] *Rakshitha Godahewa, Chang Deng, Arnaud Prouzeau, Christoph Bergmeir, 'Simulation and Optimisation of Air Conditioning Systems using Machine Learning', Researchgate, (2020) pp. 1-21.*

[4] *Robert F. Chevalier, Gerrit Hoogenboom, Ronald W. McClendon, Joel A. Paz, 'Support vector regression with reduced training sets for air temperature prediction: a comparison with artificial neural networks' Neural Computing and Applications, Springer, (February 2011) pp. 151-159.*

[5] *Y.Radhika and M.Shashi, 'Atmospheric Temperature Prediction using Support Vector Machines', International Journal of Computer Theory and Engineering, Vol. 1, No. 1, (April 2009) pp. 55-58.*

[6] *Thomas Groensfelder, Fabian Giebeler, Marco Geupel, David Schneider, Rebecca Jaeger, 'Application of machine learning procedures for mechanical system modelling: capabilities and caveats to prediction-accuracy', Advanced Modeling and Simulation in Engineering Sciences, volume 7, (2020).*

[7] *Samir Touzani, Jessica Granderson, Samuel Fernandes,' Gradient boosting machine for modeling the energy consumption of commercial buildings', Energy & Buildings, (2018) pp. 1-23.*

[8] *Subodh Paudel, Mohamed Elmitri, Stéphane Couturier, Phuong Nguyen, René Kamphuis, 'A relevant data selection method for energy consumption prediction of low energy building based on support vector machine', Energy and Buildings, (2017) pp.240-256.*

[9] *Maohui Luo, Jiaquing Xie, Yichen Yan, Zhihao Ke, Peiran Yu, Zi Wang, Jhingsi Zhang, 'Comparing machine learning algorithms in predicting thermal sensation with ASHRAE Comfort Database II', Energy & Buildings, (March 2020).*

[10] *Da Li, Carol C. Menassa, Vineet kamat, 'Robust non-intrusive interpretation of occupant thermal comfort in built environments with low-cost networked thermal cameras', Applied Energy, (Oct. 2019).*

[11] *S. Salcedo-Sanz, R. C. Deo, L. Carro-Calvo, B. Saavedra-Moreno, 'Monthly prediction of air temperature in Australia and New Zealand with machine learning algorithms', Theoretical and Applied Climatology, Volume 125, Issue 1-2, (2016) pp. 13-25.*

[12] *Yaowen Xia, Wenxian Lin, Wenfeng Gao, Tao Liu, Qiong Li, Anran Li, 'Experimental and numerical studies on indoor thermal comfort in fluid flow: A case study on primary school classrooms', Case Studies in Thermal Engineering, (2020).*