A STUDY OF THE ENERGY EFFICIENT BUILDING DESIGN TO PREDICTING HEATING AND COOLING LOADS BY ADVANCED DATA MINING APPROACH

Pham Anh Duc*, Le Thi Kim Oanh, Ho Thi Kieu Oanh

*The University of Danang, University of Science and Technology; *paduc@dut.udn.vn*

Abstract - Advanced data mining (DM) approaches are potential tools for solving civil engineering problems. This study investigates the potential use of advanced DM approaches and proposes a meta-heuristic optimization algorithm – based prediction model. This prediction model integrates the artificial firefly colony algorithm and the machine learning prediction model. The proposed model were constructed using 768 experimental datasets from the literature with 8 input and 2 output parameters, including heating load (HL) and cooling load (CL). Compared to previous works, the proposed model further obtained from at least 33.8% to 86.9% lower error rates for CL and HL prediction, respectively. This study confirms the efficiency, effectiveness, and accuracy of the proposed approach when predicting CL and HL in building design stage. Therefore, the analytical results convincedly support the feasibility of using the proposed techniques to facilitate early designs of energy conserving buildings.

Key words - cooling load; heating load; energy performance; energy-efficient building; swarm intelligence; data mining.

1. Introduction

A major challenge in many developing countries is providing sufficient energy for assisting human beings and supporting economic activities but surely minimizing any harm to society and environment. Additionally, one of energy that should be concerned is the electricity. Therefore, the social and scientific importance of electrical load forecasting has increased significantly [1]. Energy conservation is now a critical task, and buildings can achieve substantial energy savings if they are designed and operated properly. Energy awareness and management are the important measures during building lifecycle. [2]. Heating load (HL) and cooling load (CL) are used as measures of the amount of energy that must be added or removed from a space by Heating Ventilation and Air Conditioning (HVAC) system to provide the desired level of comfort within a space. Therefore, early estimations of building HL and CL can help engineers design energy-efficient buildings.

A building is considered energy-efficient if it is designed and built to decrease energy use and occupant comfort by using improved insulation, more energy-efficient windows, high efficiency space conditioning and water heating equipment, energy-efficient lighting and appliances, reduced air infiltration, and controlled mechanical ventilation. Right-sizing is selecting HVAC equipment and designing the air distribution system to achieve the expected cooling loads in the building [3]. Given current economic as well as environmental constraints on energy resources, the energy issue plays an important role in the design and operation of buildings. Therefore the best solution to alleviate the ever increasing demand for additional energy supply is to have more energy efficient building designs with improved energy conservation properties. However, accurately predicting the building heating and cooling loads is a questionable work. The Accurate load estimations have a direct impact on energy efficiency, occupant comfort, indoor air quality, and building durability. Hence, the development of models can enhance the performance's accuracy in predicting heating load and cooling load to be becoming crucial.

This study used DM approach and a meta-heuristic optimization algorithm to develop advanced data mining algorithms for solving prediction problems. The proposed advanced data mining approach integrates firefly algorithm and support vector regression (SVR) to construct an artificial firefly colony algorithm-based SVR (AFCA-SVR) model, which is a novel hybrid swarm intelligence system for forecasting problems in civil engineering. The performance of the proposed system is validated by performance comparisons with previous work *via* cross-validation algorithm and hypothesis testing.

2. Advanced Data Mining Models

Recently, researchers have raised a concern of using artificial intelligence (AI) for predicting energy consumption. Various DM techniques have been applied and proven to be a reliable and efficient tools to support energy engineers to cope with energy prediction problems. Among them, the support vector machine (SVM)-based on regression model is increasingly used in research and industry. Recent studies have used these models to analyze building energy efficiency. For example, Dong *et al.* (2005) used an SVM model to predict building energy consumption in four offices in Singapore [4]. Li *et al.* (2009) employed the SVM in regression for predicting hourly cooling demand in Guangzhou, China [5]. The SVM outperforms conventional back propagation neural networks. Hou and Lian (2009) also used SVMs to predict cooling loads in heating ventilation and air conditioning (HVAC) systems and found that SVMs are better than the autoregressive integrated moving average model [6]. Especially, Edwards *et al.* (2012) have investigated seven machine learning methods to predicting residential electrical consumption [7]. All their results showed that SVM was the best technique for predicting each home's future electrical consumption.

Several studies have attempted to develop the hybrid AI models by combining one with other technique to enhance their performance results. Hybrid computational system articles published in other civil engineering areas are also reviewed, including environmental and water resources engineering [8], highway engineering [9], and project scheduling [10]. Applications of other recent, more powerful and efficient hybrid models are also reviewed [11]. Hybrid approaches are considered a promising research area in the near future [12]. Swarm intelligence belongs to an artificial intelligence that has become a research attracted to many research scientists of related fields in recent years [13]. The swarm intelligence area has two main stages. The first stage is the ant colony algorithm [14] and the particle swarm optimization [15]. Secondly, new swarm intelligence algorithm have proposed which inspired by the behavior of honey bees [16]; fireflies [17], fish schools [18]; cuckoo birds [19]. Recently, optimization problems have been studied in both industrial and scientific contexts. Its techniques inspired by swarm intelligence have become increasingly popular [20]. Optimization problems have been studied in many fields, including tax forecasting [21], transportation engineering [22], and energy performance [23]. One optimization problem that has been studied intensively is the use of optima parameter models to improve the accuracy of the predictive results.

Briefly, the advantage of hybrid swarm intelligence approach is to use a balance trade-off between global search which is often slow and fast local searches. It is easy to combine the advantages of various algorithms so as to produce better results.

Generally, a hybrid approach made with intelligent methods will produce effective tools to solve complex problems. Therefore, this study investigates a hybrid swarm intelligence system for combining efficient AFCA with support vector machine-based regression, which can enhance the accuracy of forecasting performance in energy performance problems.

2.1. Support Vector Regression

The support vector machine developed by Vapnik in 1995 [24] has been widely used for classification, forecasting and regression. Because of their high learning capabilities, SVMs have proven effectively in the civil engineering field [4]. The SVMs can be classified into two types depending on the target: in one, the classification target has only two values (*i.e.*, 0 and 1); in the other, the regression in which the target has continuous real value. The regression model used in SVMs is SVR, a variation of an SVM for function estimation. SVR is typically used to solve nonlinear regression problems by constructing the input-output model. Typically, the regression model uses support vector regression (SVR) with a quadratic loss function, which corresponds to the conventional least squares error criterion, a variation of an SVM for function estimation to alleviate the burden of computational cost. Here, the SVR model is used to construct energy performance input-output model.

In SVR for function estimation, given a training dataset $\{x_k, y_k\}_{k=1}^{N}$, the optimization problem is formulated as Eq. (1)

$$\min_{\omega,b,e} J(\omega,e) = \frac{1}{2}\|\omega\|^2 + \frac{1}{2}C\sum_{k=1}^{N} e_k^2 \tag{1}$$

Subject to $y_k = \langle \omega, \varphi(x_k) \rangle + b + e_k, \quad k = 1,...N$

where $J(\omega,e)$ is the optimization function, ω is the parameter of the linear approximator, e_k is error variables, $C \geq 0$ is a regularization constant that specifies the constant representing the trade-off between the empirical error and the flatness of the function, x_k is input patterns, y_k is prediction labels, and N is in the sample size.

The resulting SVR model for function estimation is shown in Eq. (2):

$$f(x) = \sum_{k=1}^{N} \alpha_k K(x, x_k) + b \tag{2}$$

Where α_k, b are Lagrange multipliers and the "bias" term, respectively and $K(x,x_k)$ is the Kernel function. In highly non-linear spaces, using the kernel function in SVR as a Radial Basis Function (RBF) kernel usually yields more promising results compared to other kernels such as $K(x,x_k) = \exp(-\|x - x_k\|^2 / 2\sigma^2)$. Thus, this study applies RBF kernel functions.

2.2. Artifical Firefly Colony Algorithm (AFCA)

According to a recent literature search, the firefly algorithm, developed by Yang in 2008 [17], is very efficient and can outperform conventional algorithms such as GA and PSO in solving many optimization problems [25]. The AFCA is a stochastic, nature-inspired, meta-heuristic algorithm that can find both the global optima and the local optima simultaneously and effectively.

For a maximization problem, the brightness value can simply be set as a proportion of the value of the objective function. Other forms of brightness can be defined similarly to the fitness function in genetic algorithm. As the attractiveness of a firefly is proportional to the light intensity seen by adjacent fireflies, the attractiveness β of a firefly is defined as Eq. (3):

$$\beta = \beta_0 e^{-\gamma r^2} \tag{3}$$

Where β is the attractiveness of a firefly, β_0 is the attractiveness of a firefly at $r = 0$, r is distrance between any two fireflies, e is constant coefficient, and γ is the absorption coefficient.

The distance between any two fireflies i and j at x_i and x_j, respectively, is the Cartesian distance as presented in Eq. (4):

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^{d}(x_{i,k} - x_{j,k})^2} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{4}$$

Where r_{ij} is the distance between any two fireflies i and j at x_i and x_j, $x_{i,k}$ is the k^{th} component of spatial coordinate x_i of the i^{th} firefly, $x_{j,h}$ is the h^{th} component of spatial coordinate x_j of the j^{th} firefly, and d is search space dimension.

Equation (5) describes the movement of the i^{th} firefly when attracted to another more attractive (brighter) j^{th} firefly.

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r_{ij}^2}(x_j^t - x_i^t) + \alpha^t \theta_i^t \tag{5}$$

Where x_i^{t+1} is the coordinate of i^{th} firefly at the $(t+1)^{th}$ iteration, x_i^t is the coordinate of i^{th} firefly at the t^{th} iteration, x_j^t is the coordinate of j^{th} firefly at the t^{th} iteration, γ = absorption coefficient, which typically varies from 0.1 to 10 in most applications; β_0 = the attractiveness at $r_{ij} = 0$, α^t = a trade-off constant to determine the random behavior of movement, and θ_i^t = a vector of random numbers drawn from a Gaussian distribution or uniform distribution at time t.

This study proposes to hybridize AFCA and SVR to construct a novel artificial firefly colony algorithm based SVR system (AFCA-SVR), as novel swarm-intelligence-based algorithm to optimize SVR hyper-parameters (Fig. 1), which promotes a fast and efficient advanced model, can lead to solve real-life complex problems in civil engineering field.

Figure 1. *The hybrid artificial firefly colony algorithm based SVR system*

To automate the optimization process, AFCA was used to enable simultaneous optimization of SVR parameters. The SVR mainly address learning and curve fitting whereas the AFCA optimizes parameters C and σ to minimize prediction error. The proposed algorithm was coded in MATLAB® R2012a on a Pentium CORE 2 Quad with 2GB of RAM running Window 7. The fitness function of the AFCA was as follows:

$$f = RMSE_{Training-data} + RMSE_{Testting-data} \tag{6}$$

In the structure of the proposed model, the SVR calls the AFCA as a subroutine for optimizing its structure parameters. Thus, the objective of this model is to use the fittest SVR shapes and optimal SVR parameters to ensure acceptable estimation in optimization problems. Historical data were classified as training data and test data. The test data were used to evaluate the performance of the trained SVR model after optimization of the SVR model.

2.3. Performance Evaluation Methods

The following performance measures were used to evaluate the prediction accuracy of the proposed predictive models:

Linear Correlation Coefficient (R):

$$\mathrm{R} = \frac{n\sum y.y' - (\sum y)(\sum y')}{\sqrt{n(\sum y^2) - (\sum y)^2}\sqrt{n(\sum y'^2) - (\sum y')^2}} \tag{7}$$

Where y' is the predicted value; y is the actual value; and n is the number of data samples.

Mean Absolute Percentage Error (MAPE):

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y-y'}{y}\right| \tag{8}$$

Mean Absolute Error (MAE):

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y-y'| \tag{9}$$

Root Mean Squared Error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y'-y)^2} \tag{10}$$

Researchers often use k-fold cross-validation algorithm to minimize bias associated with the random sampling of the training and holdout data samples. Kohavi (1995) showed that ten folds are optimal (*i.e.*, ten folds obtain the shortest validation testing time acceptable bias and variance) [26].

3. The Proposed Model for Building Energy Efficiency Design

3.1. Problem Statement: Cooling and Heating Loads

Heating and cooling loads are used as the measures of the amount of energy that must be added or removed from a space by HVAC system to provide the desired level of comfort within a space. Estimating cooling and heating load is the first step of the iterative HVAC system design procedure as such Figure 2-3.

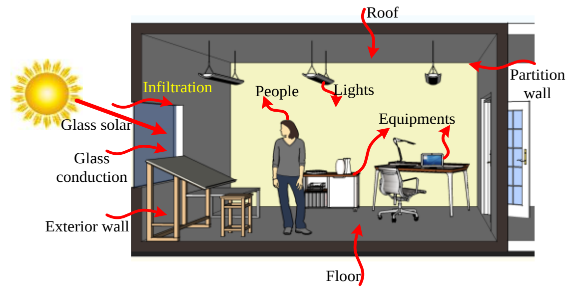

Figure 2. *The cooling load components*

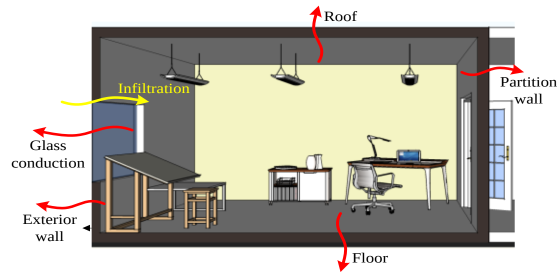

Figure 3. *The heating load components*

3.2. Data description and preparation

Heating and cooling loads in the building is affected by many parameters, which can be grouped into two main categories: the optical and thermal properties of building and the meteorological data. In this case, the dataset includes eight input variables (*i.e.*, relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area, and glazing area distribution) and two output variables (heating load and cooling load), which were simulated by Tsanas and Xifara (2012) [27]. The dataset comprises 768 samples and 8 features, aiming to predict two real valued responses (Table 1). These variables have been frequently used in the energy performance of building literature to study energy related topics in buildings [28].

Table 1. *Statistical parameters for energy performance of building*

Parameters	Unit	Min.	Ave.	Max.
Relative compactness	N/A	0.62	0.76	0.98
Surface area	m^2	514.50	671.71	808.50
Wall area	m^2	245.00	318.50	416.50
Roof area	m^2	110.25	176.60	220.50
Overall height	m	3.50	5.25	7.00
Orientation	N/A	2.00	3.50	5.00
Glazing area	%	0.00	23.0	40.0
Glazing area distribution	N/A	0.00	2.81	5.00
Heating load	kW	6.01	22.31	43.10
Cooling load	kW	10.90	24.59	48.03

4. Results and discussion

In this study, k-fold cross validation method is used to ensure good generalization capability. The performance of the proposed prediction model is validated in terms of R, RMSE, MAE and MAPE. A high R value and low RMSE, MAE and MAPE values indicate good performance of the model. Table 2 presents the improvement and hypothesis testing of the AFCA-SVR models *via* cross-fold validation algorithm. Tsanas and Xifara (2012) proposed a classical linear regression approach as iteratively reweighted least squares (IRLS) and classification using random forests (RF) to estimate heating load (HL) and cooling load (CL) [27], their results obtained 10.09%, 2.18% and 9.41%, 4.61% for MAPE in HL and CL cases, respectively.

In the classic linear regression approach used to estimate heating load [27], IRLS and RF models obtained MAPEs of 10.09% and 2.18%, respectively in heating load cases (Table 2). The AFCA-SVR model obtained a lower MAPE (1.43%). It also was lower in MAE (0.29 kW) for heating load case compared to the IRLS, RF models (2.14 kW and 0.51 kW, respectively). Overall, error rates improved by AFCA-SVR model were 34.2%–86.9% compared to those of previous models in heating load cases. The hypothesis testing results confirmed the significantly improved performance of the AFCA-SVR model at 1% of the α level by their *p-values*.

Table 2. *Hypothesis testing results and improvement rates in the AFCA-SVR model*

Empirical models	Performance measure					% improved by AFCA-SVR			
	R (%)	RMSE (kW)	MAE (kW)	MAPE (%)		R	RMSE	MAE	MAPE
Heating load									
IRLS	N/A	3.14	2.14	10.09		N/A	86.9*	86.6*	85.8*
RF	N/A	1.01	0.51	2.18		N/A	59.2*	43.7*	34.2*
AFCA-SVR	99.9	0.41	0.29	1.43					
Cooling load									
IRLS	N/A	3.39	2.21	9.41		N/A	54.3*	57.7*	76.1*
RF	N/A	2.57	1.42	4.62		N/A	39.7*	33.8*	51.3*
AFCA-SVR	98.0	1.55	0.94	2.25					

Note: *The improvement and hypothesis testing are calculated using average performance measures;*
** indicates significance level is higher than 1%*

The tests yielded statistically significant results at 1% of the α level by their *p-values*, rejecting the null hypothesis (*i.e.*, modeling performance of previous works equaled or exceeded the results of the AFCA-SVR model). The hypothesis tests verify that performance measures were significantly improved for the AFCA-SVR model. For example, for a linear correlation coefficient of R = 98.0%, AFCA-SVR model obtained a lower MAE (0.94 kW) for cooling load case compared to that (2.21 kW; 1.42 kW) of IRLS and RF models, respectively. Overall, the percentage of the error rates improved by the AFCA-SVR model were 33.8%–76.1% lower than those of previous models in cooling load cases. The simulated values provided by Ecotect for HL and CL are considered to reflect the true actual values. However, a detailed comparison of the provided output values from different simulation package is beyond the scope of this case.

5. Conclusions

The proposed approach is performed and has many potential applications in building energy prediction. Various building characteristics were used as input to HL and CL Data for 768 cases of CL and HL that were used to construct the prediction models. A 10-fold cross-validation method was used to mitigate the bias in comparisons of the model performance. The analytical results demonstrate the applicability of advanced data-mining technique for forecasting energy consumption by buildings. The civil engineering problems are inherently heterogeneous and enormously complex. It is also influenced by highly variable and unpredictable factors. Because of these difficulties and the importance of enhancing estimation capability, the complexity approaches (integrated models) have been used to develop algorithms that improve modeling accuracy, effectiveness, and speed.

Recognizing the need for effective trade-off tools and the potential drawback of state-of-art predictive model, the main purpose of this study is to establish a hybrid swarm intelligence system. This system is named the hybrid artificial firefly colony algorithm-based SVR model that can solve effectively forecasting problems in building energy performance. In the cooling and heating load prediction, the experimental results have demonstrated that AFCA-SVR can achieve more than 33.8% reduction in prediction error rates compared to other benchmark methods.

Future studies may also evaluate the use of the proposed approach for automatic parameter tuning and efficient improvement on civil engineering and management. For example, since the environmental sustainability is now a very important global issue, future buildings must be highly energy efficient without compromising the comfort and safety of occupants. This study confirms that the proposed swarm intelligence-based prediction model can assist building owners, facility managers, operators, and tenants of buildings in assessing, benchmarking, diagnosing, tracking, forecasting, and simulating energy consumption in building portfolios.

REFERENCES

[1] Ahmed A, Korres NE, Ploennigs J, Elhadi H, Menzel K. Mining building performance data for energy-efficient operation. *Advanced Engineering Informatics*, 2011;25(2):341-54.

[2] Al-Homoud MS. Computer-aided building energy analysis techniques, *Building and Environment*, 2001;36(4):421-33.

[3] Burdick A. Strategy Guideline: Accurate Heating and Cooling Load Calculations, *US Department of Energy: IBACOS*, Inc.; 2011.

[4] Dong B, Cao C, Lee SE. Applying support vector machines to predict building energy consumption in tropical region, *Energy and Buildings*, 2005;37(5):545-53.

[5] Li Q, Meng Q, Cai J, Yoshino H, Mochida A. Applying support vector machine to predict hourly cooling load in the building, *Applied Energy,* 2009;86(10):2249-56.

[6] Hou Z, Lian Z. An Application of Support Vector Machines in Cooling Load Prediction, *International Workshop on Intelligent Systems and Applications, ISA 20092009*. p. 1-4.

[7] Edwards RE, New J, Parker LE. Predicting future hourly residential electrical consumption: A machine learning case study, *Energy and Buildings*, 2012;49(0):591-603.

[8] Li L, Tian G. An improve wavelet relevance vector regression algorithm for water resource demand prediction, *Journal of Information & Computational Science,* 2012;9(1):183–9.

[9] Callow D, Lee J, Blumenstein M, Guan H, Loo Y-C. Development of hybrid optimisation method for Artificial Intelligence based bridge deterioration model — Feasibility study, *Automation in Construction,* 2013;31(0):83-91.

[10] Chen P-H, Shahandashti SM. Hybrid of genetic algorithm and simulated annealing for multiple project scheduling with multiple resource constraints, *Automation in Construction,* 2009;18(4):434-43.

[11] Hassanzadeh T, Meybodi MR. A new hybrid approach for data clustering using firefly algorithm and K-means, *Artificial Intelligence and Signal Processing (AISP), 2012 16th CSI International Symposium on 2012,* p. 007-11.

[12] Foucquier A, Robert S, Suard F, Stéphan L, Jay A. State of the art in building modelling and energy performances prediction: A review, *Renewable and Sustainable Energy Reviews,* 2013; 23(0):272-88.

[13] Yang X-S, Cui Z, Xiao R, Gandomi AH, Karamanoglu M. Swarm *Intelligence and Bio-inspired Computation - Theory and Applications,* Waltham, MA 02451, USA: Elsevier Inc.; 2013.

[14] Dorigo M, Caro GD. The ant colony optimization meta-heuristic. In: David C, Marco D, Fred G, Dipankar D, Pablo M, Riccardo P, et al., editors. *New ideas in optimization*: McGraw-Hill Ltd., UK; 1999. p. 11-32.

[15] Kennedy J, Eberhart R. Particle swarm optimization. Neural Networks, 1995 Proceedings, *IEEE International Conference on 1995,* p. 1942-8 vol.4.

[16] Abbass HA. MBO: marriage in honey bees optimization-a Haplometrosis polygynous swarming approach. *Evolutionary Computation, 2001 Proceedings of the 2001 Congress on 2001*. p. 207-14 vol. 1.

[17] Yang X-S. *Firefly algorithm,* Bristol, UK: Luniver Press; 2008.

[18] Cai Y. Artificial Fish School Algorithm Applied in a Combinatorial Optimization Problem, *International Journal of Intelligent Systems and Applications: MECS; 2010,* p. 37-43.

[19] Yang X-S, Deb S. Cuckoo Search via Lévy flights. *World Congress on Nature & Biologically Inspired Computing,* NaBIC 20092009. p. 210-4.

[20] Yang X-S. *Engineering optimization: an introduction with metaheuristic applications,* UK: John Wiley & Sons; 2010.

[21] Liu L-X, Zhuang Y-Q, Liu X-Y. Tax forecasting theory and model based on SVM optimized by PSO, *Expert Systems with Applications,* 2011;38(1):116-20.

[22] Chou J-S, Le T-S. Reliability-based performance simulation for optimized pavement maintenance, *Reliability Engineering & System Safety,* 2011;96(10):1402-10.

[23] Coelho LdS, Mariani VC. Improved firefly algorithm approach applied to chiller loading for energy conservation, *Energy and Buildings,* 2013;59(0):273-8.

[24] Vapnik VN. *The nature of statistical learning theory,* New York: Springer-Verlag; 1995.

[25] Pal SK, Rai CS, Singh AP. Comparative Study of Firefly Algorithm and Particle Swarm Optimization for Noisy Non-Linear Optimization Problems, *International Journal of Intelligent Systems and Applications 2012,* p. 50-7.

[26] Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection Paper presented at *the International Joint Conference on Artificial Intelligence*. 1995.

[27] Tsanas A, Xifara A. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools, *Energy and Buildings,* 2012;49(0):560-7.

[28] Pacheco R, Ordóñez J, Martínez G. Energy efficient design of building: A review, *Renewable and Sustainable Energy Reviews,* 2012;16(6):3559-73.

(The Board of Editors received the paper on 26/10/2014, its review was completed on 31/10/2014)