Proceedings of International Joint Conference on Neural Networks, Dallas, Texas, USA, August 4-9, 2013

Multi-Pattern Cross Training: An ANN model training method using WSN sensor data

Yi Zhao, Valentin Gies, Ademir Felipe Teles, Jean Marc Ginoux

***Abstract*— A wireless sensor network (WSN) consisting of autonomous sensor nodes can provide a rich stream of sensor data representing physical measurements. A well built Artificial Neural Network (ANN) model needs sufficient training data sources. This paper proposes a procedure of combining ANN and WSN sensor data in modeling. Experiments on indoor thermal modeling demonstrated that WSN together with ANN can lead to accurate fine grained thermal models. A new training method "Multi-Pattern Cross Training"(MPCT) is also introduced in this work. This training method makes it possible to merge knowledge from different training data sources (patterns) into a single ANN model. Further experiments demonstrated that models trained by MPCT method shew better generalization performance and lower prediction errors in tests using different data sets.**

I. Introduction

WSN offers a practical solution of distributed sensing, processing, communication and control. Generally, WSN could be described as a network of distributed self-powered nodes that could sense or exchange with environment. The main advantage of WSN is that it could be easily and rapidly installed and gather information for a long period of time, providing an enormous quantity of sensor data. WSN based applications have shown a rapid growth in a variety of fields, including target tracking and surveillance, natural disaster relief, health monitoring, environment exploration and geological sensing[1].

One of the most common applications of ANN is to model the behavior of certain system. ANN's self-adaptivity and nonlinear mapping ability make it more advantageous in modeling nonlinear system or system with unknown dynamics[2][3].

So far, a modest number of researches have combined WSN with ANN in modeling. We think the combination of WSN and ANN can be a powerful modeling solution. First, A trainable ANN model built itself from experimental data, thus, sufficient data sources are necessary to obtain an accurate ANN model. The rich sensor data from WSN in return can be used to meet this need. Second, WSN data based ANN modeling has high practical values: The behavior of certain system is very complex and difficult to analyze, especially when many nonlinear and time-varying effects are present, for example, the dynamic behavior of building in response to environmental factors. In such case, it is nearly impossible to obtain an accurate mathematical model with limited system parameters. Thus, an adaptive ANN thermal model using WSN sensor data can be a reasonable choice. Our proposed modeling solution is presented below in Fig. 1.

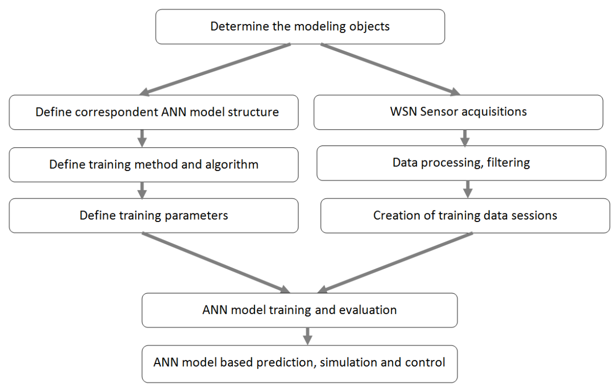


Fig. 1. The WSN data based ANN modeling procedure.

In our recent research on building thermal modeling, we combined ANN and WSN sensor data to establish an indoor thermal model. A practical problem we have is this: On one hand, long term WSN data acquisition causes high data redundancy which leads to low training efficiency and high computational cost in modeling, on the other hand, Short acquisition with limited duration is usually obtained under certain specific conditions. Because environmental conditions change greatly, Short acquisition as a training source only contains limited information. The result is that the trained ANN model has poor generalization performance. The model shows low prediction error against its training data, but relatively higher predictions error against other test data measured under different conditions. So, the question we raised here is "Is it possible to use a single ANN model to obtain a more complete understanding on the system's behavior?"

Indeed, there are existing well-known methods developed for training neural networks[4][5], however, they are not aimed to build a more comprehensive ANN model. Perturbation method and sensitivity analyzing[6] can be used to improve the training efficiency by reducing the redundancy as well as the input layer dimension. This method is very useful when the some of the network's inputs are correlated.

The term "Cross Training" was originally defined as a

Yi Zhao ,Valentin Gies and Jean Marc Ginoux are with the Laboratory PROTEE, EA3819, University of the South, Toulon Var, La Garde, 83130, France,(email: {yzhao,vgies, ginoux}@univ-tln.fr).

Ademir Felipe Teles is a current graduate student in ENSTA Paris-Tech, 828, Boulevard des Marechaux, Palaiseau, 91762, France,(email: afelipet@gmail.com)

978-1-4673-6129-3/13/$31.00 ©2013 IEEE

training method for athletes to improve their competitive performance in a certain sport by systematically training in a variety of different sports[7]. Later, Gökhan H.Bakır has introduced this concept into his research on SVMs training[8].

Here, we plant this concept into a new ANN model training method "Multi-Pattern Cross Training"(MPCT). Using the internal properties of neural networks, this training method is capable of merging knowledge from different training data patterns into a single network model. Thus, by exploring the generality in the system behaviors, it can adequately be used to describe a more complete phenomena. Our ANN modeling methodology is presented in Fig. 2, while the MPCT method is highlighted in the bold black frame in the right part of Fig. 2.

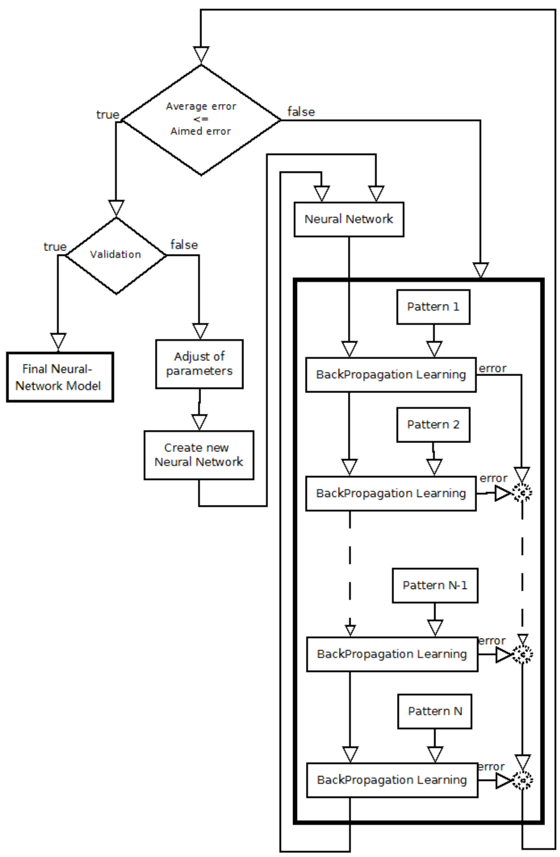
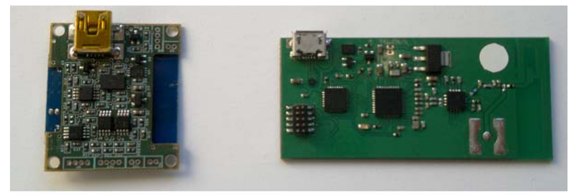


Fig. 2. Multi-Pattern Cross Training Method.

II. WSN Sensor data acquisition and MPCT method integrated software design

Industrial grade hardware and software solutions are developed in this work, including a WSN based sensor acquisition system and an ANN integrated Graphic User Interface (GUI) software. The WSN are based on (TI) CC2530 microcontroller. It contains Zigbee 2007 specification and supports both Zigbee and Zigbee PRO feature sets. [1]The WSN system as a typical mesh network contains three types of module: Coordinator, Router, and Enddevice. Different type of Sensors are integrated on the Enddevices (see Fig. 3). The sensor data are transferred through RF signal to the Coordinator. The Coordinator then gathers the sensor data and send them to PC.



End-Device(Sensor Node) Coordinator

Fig. 3. Enddevice and Coordinator of the WSN system.

The performance of our latest WSN hardware platform are verified. Its technical parameters are summarized in Tab. I.

TABLE I
Parameters and features of the WSN system.

Feature of the WSN system	Values
Cost per unit	8 euros
Network Coverage	20m/110m
Consumption in operational mode	40mA
Consumption in Power saving mode	126μA
Battery Life	>200h
Reliability	High
Packet loss ratio	<0.02
Installation	Easy
Network capacity(Nodes per Network)	25920

The main features of this WSN system are low cost, low consumption, high reliability and high network capacity.

The ANN integrated GUI software are developed under Visual Studio. The sensor data are transferred via USB to the software. The software's four main functions are data storage, signal processing, ANN modeling, simulation and prediction.

The modeling object in reality can usually be described as a Multi-Input-Multi-Output(MIMO) system. Accordingly, we defined a Three-layered back-propagation neural network(BPNN) structure to fit this MIMO system. BPNNs, proposed by Rumelhart *et al.* [9], is one of the most commonly used neural network structures. Hecht-Nielsen demonstrated that a three-layered BPNN is capable of approximating continuous mapping[10]. We also integrated both basic back-propagation learning and MPCT method in the software to evaluate their different modeling performance. As we

[1]ZigBee is a low-cost, low-power, wireless mesh network specification based on IEEE802.15.4. It is widely deployed in wireless control and monitoring applications.

explained in introduction, in some cases, the system behavior of modeling object is very complex, specially when it contains lots of nonlinear and time varying effects. If we build an ANN thermal model using single session of acquisition (single training pattern) as training source, the trained network model usually shows high prediction errors mainly because the acquisition is insufficient to cover the complete system behavior. An approach to addressing this issue is to use MPCT method to train a more comprehensive neural network model using different acquisitions as the training source.

The mechanism of MPCT method is presented below: The training source consists of WSN sensor data which is called a session of acquisition (pattern) $S_{(1)}$ [2]. Acquisitions under different initial conditions have been made, so we have different sessions of acquisitions "$S_{(p)}$". The error function E is defined as the sum of squares on the differences between the WSN sensor measured data (measured output) and the network output y_j. If we use on-line learning mode, the error obtained from the data set n of pattern $S_{(1)}$ is:

$$E_{S_{(1)}}(n) = \frac{1}{2}\sum_{j=1}^{j=N_L}(S_{(1)}(j) - y_j)^2 \tag{1}$$

j is one output layer neuron, N_L is the number of neurons in the output layer and n is the sequence number of data set in a session of acquisition. Then,the weight update value $\Delta\omega_{ij}(n)$ as:

$$\Delta\omega_{ij(S_{(1)})}(n) = -\eta\frac{\partial E_{S_{(1)}}(n)}{\partial\omega_{ij}(n)} \tag{2}$$

η is the learning rate which represents the step size on this gradient direction. At the end of this training epoch, instead of recycling the weight update with pattern $S_{(1)}$, we lead the training procedure to the training data sequence 1 in pattern $S_{(2)}$. Here, we define the last data pattern sequence of session 1 is N_1. We also introduce the momentum term to incorporate the past weight updates into the present weight update. Thus, the new weight update at the beginning of the second training epoch can be put as:

$$\Delta\omega_{ij(S_{(2)})}(n) = -(1-\alpha)\eta\frac{\partial E_{S_{(2)}}(1)}{\partial\omega_{ij}(n)} + \alpha\Delta\omega_{ij(S_{(1)})}(N_1) \tag{3}$$

where α is the momentum parameter which determines the amount of influence from the previous weight update. n refers to the last data pattern sequence in session1.

So If we define the two main variable in the equation below (Eq.4 Eq.5)as the session number p, and the sequence number n while the total session number is P and the last sequence count for session p is N_p, we can thus translate the MPCT method into computational language below:

If we define a nested loop with one outer loop and one inner loop:

The Outer loop is : for(p=1; p<=P; p++).

The Inner loop is : for(n=1; n<=N_p; n++).

The general expression of MPCT method can be put as:

$$\begin{aligned} &Outerloop(Innerloop(\Delta\omega_{ij(S_{(p)})}(n) = \\ &-(1-\alpha)\eta\frac{\partial E_{S_{(p)}}(n)}{\partial\omega_{ij}(n)} + \alpha\Delta\omega_{ij(S_{(p)})}(n-1))) \end{aligned} \tag{4}$$

Specially, when one training epoch is completed and the next epoch begins with the next data session, the weight update is equal to:

$$\Delta\omega_{ij(S_{(p)})}(n) = -(1-\alpha)\eta\frac{\partial E_{S_{(p)}}(n)}{\partial\omega_{ij}(n)} + \alpha\Delta\omega_{ij(S_{(p-1)})}(N_{(p-1)}) \tag{5}$$

The training procedure continues until the network error is less than the target error which is considered sufficiently small.

To achieve expected modeling results with MPCT method, some preparations are required. Initially, each training data session should be verified so that they contains well-defined information of system behaviors. At the same time, the neural network should be capable of recognizing these patterns' behaviors. Secondly, it is essential to ensure that the neural network could discern between different sessions of acquisition. Therefore, it is necessary to have at least one input containing a particular value for each session. This value allows the network to distinguish between the difference in patterns, thereby preventing the system from converge into one intermediary solution. One special point that has to be made here: in order to make the ANN model more practical, the selecting of data session should respect the phenomena's basic statistical distribution. For example, in our work of indoor thermal modeling, two inputs are based on the outdoor condition. As our experiments are carried out in the south of France, the presence of measured environmental data session should obey with local Mediterranean climate statistical characteristics. For example, data sessions obtained under extreme weather conditions should be avoided in the training data source.

As for the definition of basic network training parameters, we give two possibilities in this software to chose these parameters: 1. Manual selection, the parameters can be defined by users as presented in Fig.4.

2. Auto selection, the parameters can selected by a cyclic algorithm which is also presented in Fig.2: a different range of training parameters are used to train network models, the trained models' prediction errors regarding its training data are then automatically calculated. In the end, the software will select the parameter combination which lead to minimum prediction errors.

III. ANN THERMAL MODEL: CONCEPTION AND EXPERIMENTS

Research on thermal modeling brings two main interests: First, amelioration of indoor condition and human comfort, improving storage conditions for thermal-sensitive products. Second, a precis indoor thermal model is necessary for an optimum Heating/Cooling strategy which can

[2]A session of acquisition consists of sets of time-series sensor data collected during a certain period

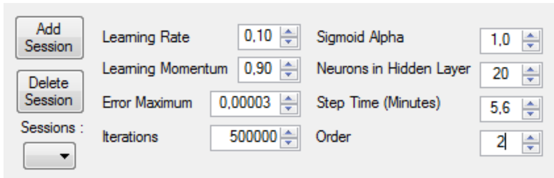


Fig. 4. Define training parameters manually.

eventually contribute to high energy efficiency and lower consumptions[11]. In this research, we try to establish a fine grained building room thermal model by combining ANN and WSN sensor data. The first step is to define the modeling object: the model's inputs and outputs. In our case, the outputs are the indoor temperatures in each zone of the room since indoor temperature is one of the most important factors in indoor environment. The model's main input is the operating status of indoor heating source while the disturbance inputs are the environmental factors such as Solar radiation, outdoor temperature. Since there is no absolute calculation law on the number neurons in the hidden layers, we estimated it with the formula below:

$$Num.\ of\ Hidden\ Neurons = \frac{1}{2}(N_i + N_o) + \sqrt{N_{(tp)}} \quad (6)$$

N_i is the dimension of network inputs, N_o is the dimension of outputs, $N_{(tp)}$ is the total number of training data patterns. this formula has been used in several engineering problems for modeling and prediction with good results[12][13]. The inputs, outputs and hidden layer together frame a three-layered BPNN structure. In order to increase the model's accuracy, we also proposed higher order models by including the previous output temperature t-Ti in the input layer (see Fig. 5), or even t-Ti and t-2Ti both in the input layer.

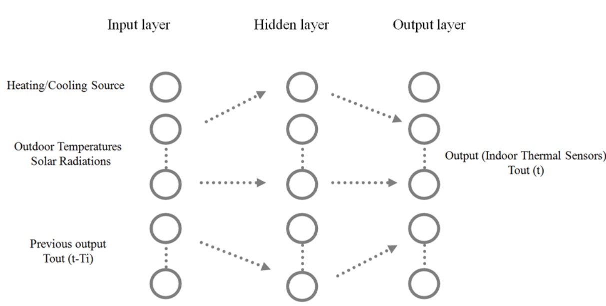


Fig. 5. A second order ANN thermal model structure.

A sigmoid function (see Eq.7) is chosen as activation function $f(x)$ in the software, because its nonlinear characteristic makes it suitable for modeling the indoor temperatures changes.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (7)$$

Experiments have been carried out in an office room (Width 5.5m, Length 7.2m, Height 3.15m) located in the building E of IUT in the campus of University of the South, Toulon Var (LAT 43°123′ N, LONG 6°11′ E), city of La Garde, south France. Three temperature sensors placed outside the building room, one infrared sensor, and the indoor Heating/Cooling unit's operating status provide the data as network inputs. Five thermal sensors as network outputs are placed horizontally at the height of 1.10m in the room where most human activities take place[3](see Fig.6). We carried out the WSN sensor acquisitions under different environmental conditions during four month from the Jun to September 2012.

The aim of the thermal modeling is to find the thermal response(Heating/Cooling)of the building room under different indoor and outdoor conditions. The sessions of acquisition

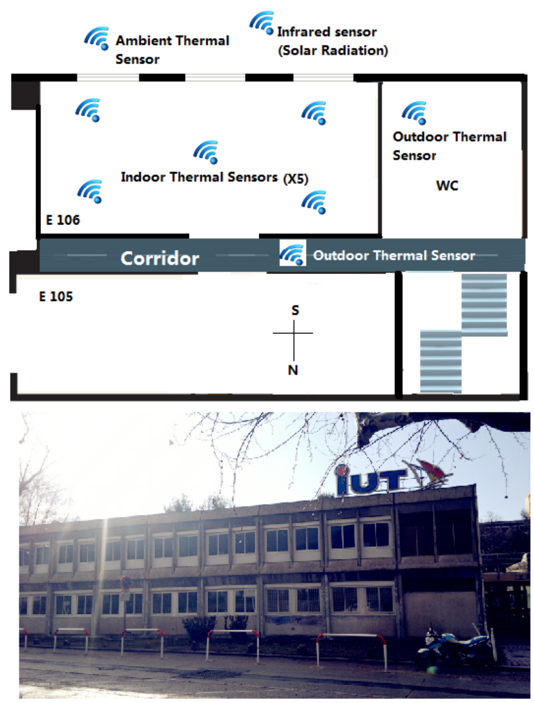


Fig. 6. Experiments in Building E of University of the south, Toulon Var. France

are stored according to their time sequence in the software's database. Using these sensor data, we trained two types of network models: one type trained with a single session of sensor data and the other trained by the MPCT method with two to four different data sessions. These four sessions are representative because they are all measured under different environmental conditions with a minimum time interval of 10 days. The test data sessions are also measured during the same period under different conditions.

[3]According to air-conditioning industries in France, the room temperature is usually evaluated at the height of 1.10m

IV. Modeling results and discussions

In order to evaluate the modeling results, an Average Mean Squared Error (AMSE) is calculated as below (see Eq.(7)-(8)). If we consider the model predicted temperature as T', the measured temperature as T, the number of measures as n; the Mean Squared Error (MSE) for one output (one sensor) is:

$$MSE = \frac{1}{n}\sum_{i=2}^{n+1}\left(T^{'}(i) - T(i)\right)^{2} \tag{8}$$

the AMSE is the average MSE value of all the k outputs (k sensors):

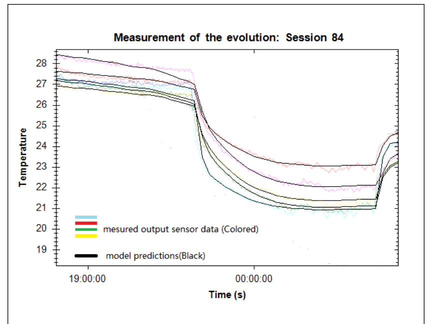
$$AMSE = \frac{1}{k}\sum_{i=1}^{k} MSE(k) \tag{9}$$

At first, the thermal model is trained with WSN sensor data from a single session. we compared its predictions errors regarding its training data session and test data session No.1 which is measured under similar outdoor and indoor conditions. The results are presented in Tab.II We repeated similar acquisitions and the modeling results

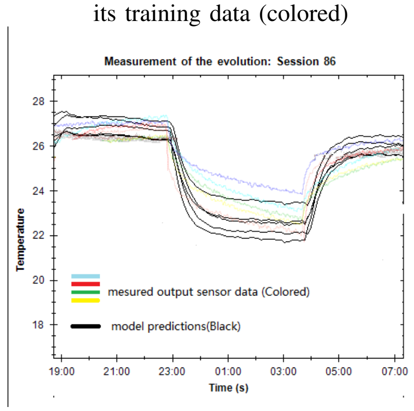
TABLE II
AMSE of single pattern trained ANN thermal model regarding training and similar test data

Performance Criteria	AMSE
Against Training data	0.1179°C
Against Test data	0.2142°C

positively supported our assumption in the introduction. It shows that WSN sensor data and ANN together can bring accurate fine grained indoor thermal models. The model characterized the thermal response on different points of a building room; predicted temperature drops is consistent with the real measurement. The model's prediction error was very small compared to the total indoor temperature variation. The results also suggest that the ASME of the single pattern trained model is acceptable when it is applied to similar test data. However, following experiments found that its performance is not so convencing when facing test data which is measured under different indoor and outdoor conditions. We compared the model response regarding its training data session 84 in Fig. 7a, with that regarding test data session 86 in Fig. 7b. The originally measured outputs (multiple sensors) are colored, the model responses are in black.
It is apparent in Fig.7 that a higher model prediction error is presented when applying the model to testing data measured under different conditions. The main reason, as explained in introduction, is that the single session of WSN acquisition as training data contains limited information, so the trained network model cannot cover the whole phenomena.
The modeling results using MPCT method are presented below in Fig. 8. Session 84 is one of the three training



a. Single pattern trained model response (black) regarding its training data (colored)

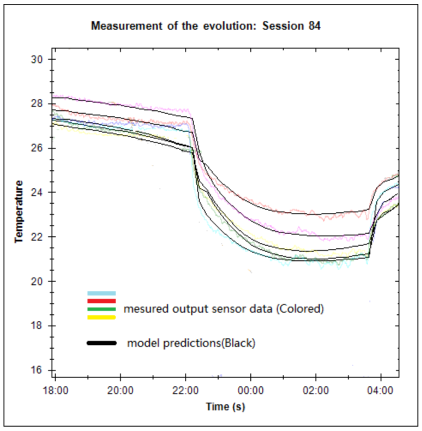


b. Single pattern trained model response (black) regarding measured test data (colored)

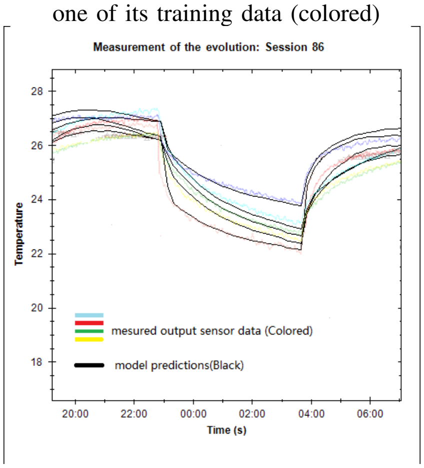
Fig. 7. Model trained with single pattern

sessions used for this MPCT model while session 86 is a test data session which is not used in training the model. The originally measured outputs (multiple sensors) are colored, the model responses are in black.

We found that models using Pattern Cross training Method have much lower prediction errors regarding test data. The reason is that with MPCT method, the trained model can merge knowledge from different training sources, understand and cover a wider system behaviors, thus, its generalization performance outperformed the model trained with single session.
Another particular argument for using the MPCT method should also be discussed here. Based on the back-propagation algorithm, the MPCT method was used in successive training epochs while alternating between the separated data sessions. For each epoch, the neural network tends to adapt the model for the present data session. Thus, by changing the input data session, the convergence direction changes too. In this way, the presence of local minima can probably be avoided.
More experiments allowed us to further compare their different modeling results in Tab.III. Trn pattern refers to the training data pattern while Tst Pattern refers to Test data pattern. These experiments confirm again that models trained with the MPCT method have lower prediction error on differ-



a. MPCT model response (black) regarding one of its training data (colored)



b. MPCT model response (black) regarding measured test data (colored)

Fig. 8. Model trained with MPCT method

ent test data. However, it indicates that the increase of data sessions used in MPCT training does not guarantee lower prediction errors. As we have mentioned in the introduction of this paper, the MPCT modeling quality is directly related to the capacity of the network and the structuring of training data sessions. Limited differences or data redundancy in the selected data sessions may cause low efficiency of MPCT training.

V. Conclusion

This paper has demonstrateed that the combination of WSN and ANN is a suitable solution in thermal modeling. It also presents an ANN model training method "Multi-Pattern Cross Training"(MPCT). The training method makes it possible to train a single ANN model with multiple training data sessions. Experiments of WSN sensor data based ANN thermal modeling suggest that the MPCT trained model outperforms single pattern trained model on its generalization performance. A model trained with the MPCT method shows lower prediction errors. The potential of "Cross Training" has not yet been fully discussed in this paper. The ultimate goal of MPCT method is to use minimum training data sessions to obtain a model with best generalization performance. To realize this, further exploration can be undertaken in the following aspect: to find the optimum match between neural network capacity and the structuring of training data sessions.

TABLE III
Training and Testing comparison between Single Pattern training and MPCT training

Data	AMSE (single pattern model)	AMSE (MPCT trained models)		
Data Pattern	Single pattern	2 patterns	3 patterns	4 patterns
Trn Pattern 1	0.1179	0.1292	0.131	0.1194
Trn Pattern 2	-	0.1479	0.131	0.1394
Trn Pattern 3	-	0.1214	0.1105	0.1240
Trn Pattern 4	-	0.1501	0.1571	0.1724
Tst Pattern 1	0.2142	0.212	0.1982	0.2038
Tst Pattern 2	0.9571	0.2720	0.2589	0.2990
Tst Pattern 3	1.4172	0.2435	0.2724	0.2865
Tst Pattern 4	2.2414	0.3675	0.2959	0.3745
Tst Pattern 5	2.5720	0.2514	0.3720	0.2984

References

[1] Jennifer Yick, Biswanath Mukherjee, and Dipak Ghosal, "Wireless sensor network survey," *Computer networks*, vol. 52, no. 12, pp. 2292–2330, 2008.
[2] MY Rafiq, G Bugmann, and DJ Easterbrook, "Neural network design for engineering applications," *Computers & Structures*, vol. 79, no. 17, pp. 1541–1552, 2001.
[3] Kumpati S Narendra and Kannan Parthasarathy, "Identification and control of dynamical systems using neural networks," *Neural Networks, IEEE Transactions on*, vol. 1, no. 1, pp. 4–27, 1990.
[4] Hung-Han Chen, Michael T Manry, and Hema Chandrasekaran, "A neural network training algorithm utilizing multiple sets of linear equations," *Neurocomputing*, vol. 25, no. 1, pp. 55–72, 1999.
[5] Robert S Scalero and Nazif Tepedelenlioglu, "A fast new algorithm for training feedforward neural networks," *Signal Processing, IEEE Transactions on*, vol. 40, no. 1, pp. 202–210, 1992.
[6] Jacek M Zurada, Aleksander Malinowski, and Shiro Usui, "Perturbation method for deleting redundant inputs of perceptron networks," *Neurocomputing*, vol. 14, no. 2, pp. 177–193, 1997.
[7] Hirofumi Tanaka, "Effects of cross-training," *Sports Med*, vol. 18, no. 5, pp. I994, 1985.
[8] Gökhan Bakır, Léon Bottou, and Jason Weston, "Breaking svm complexity with cross training," *Advances in neural information processing systems*, vol. 17, pp. 81–88, 2005.
[9] David E Rumelhart, Geoffrey E Hintont, and Ronald J Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
[10] Robert Hecht-Nielsen, "Theory of the backpropagation neural network," in *Neural Networks, 1989. IJCNN., International Joint Conference on*. IEEE, 1989, pp. 593–605.
[11] K.H.YANG and C.H.SU, "An approach to building energy savings using the pmv index," *Building and environment*, vol. 32, pp. 25–31, 1997.
[12] Soteris A Kalogirou, Constantinos C Neocleous, and Christos N Schizas, "Artificial neural networks for modelling the starting-up of a solar steam-generator," *Applied Energy*, vol. 60, no. 2, pp. 89–100, 1998.
[13] Soteris A Kalogirou and Milorad Bojic, "Artificial neural networks for the prediction of the energy consumption of a passive solar building," *Energy*, vol. 25, no. 5, pp. 479–491, 2000.