ESTIMATING THE COST OF ENERGY USAGE IN SPORT CENTRES: A COMPARATIVE MODELLING APPROACH

A.H. Boussabaine, R.J. Kirkham and R.G. Grew

Construction Cost Engineering Research Group, School of Architecture and Building Engineering, University of Liverpool, Leverhulme Building, Liverpool L69 3BX.[1]

The cost of provision and maintenance of safe and comfortable conditions for the activity taking place in any leisure centre are the predominant factors in the prediction of running costs. This paper describes the fundamentals associated with running costs in leisure centres and then investigates nineteen sport centres in the City of Liverpool, United Kingdom using data elicited from the Liverpool Leisure Services Directorate. The energy operating costs were analysed using statistical techniques and artificial intelligence methods. Three types of modelling, linear / non-linear regression, neural networks and neuro-fuzzy were developed to predict total energy cost. Testing and validation of the results showed that neural network models outperformed both regression and neuro-fuzzy techniques. However, all the models showed a high level of accuracy. The models would be of use to professionals involved in feasibility studies of different scenarios at the design stage.

Keywords: artificial intelligence, cost modelling, running cost, sport centre, statistics.

INTRODUCTION

The cost of energy has become an increasingly important consideration at both design and handover stage of any building. However, some buildings have inherently higher energy demands than others, such as sport centres, and it is here that effective energy cost modelling can be used to make accurate forecasts of likely consumption patterns. Deciding through which type of building to include in a forecasting model is not the only problem. The choice of modelling technique is also important. Statistical models have been used for some time but here, artificial intelligence is proposed as a more reliable and accurate modelling technique. Providing professionals with accurate forecasting techniques will enable them to make educated and reliable estimates of likely energy consumption in sport centres, as well as other forms of building.

THE PROBLEM STATEMENT

The task of forecasting energy usage can be posed as a classification problem, given a set of classes and a set of input data instances, each described by a suitable set of features, assign each input data instance to one of the classes. For this paper, the different characteristics of sport centres form the set of input data instances and the various energy costs form the set of possible classes to which the input characteristic

1 a.h.boussabaine@liv.ac.uk, r.j.kirkham@liv.ac.uk, r.g.grew@liv.ac.uk
http://www.liv.ac.uk/~halim/concostgroup.html

Boussabaine, A H, Kirkham, R J and Grew, R G (1999) Estimating the cost of energy usage in sport centres: a comparative modelling approach. *In:* Hughes, W (Ed.), *15th Annual ARCOM Conference*, 15-17 September 1999, Liverpool John Moores University. Association of Researchers in Construction Management, Vol. 2, 481-8.

of sport centres can belong. Each sport centre can be described by a set of features, which represent important information about the usage of energy.

Let ***p*** represent the space of ***n*** sport centres therefore: $P_1, P_{2\ldots\ldots}P_n$, and r be the set of possible (mutually exclusive) m classes of sport centre ratings according to the cost of energy usage, $R_1, R_{2\ldots\ldots}R_n$. Let ***f*** represent the k dimensional feature space $F_1, F_{2\ldots\ldots}F_n$ describing each sport centre. Each sport centre can be considered as a k-cluster $F_{1_{Bi}}\cdots\cdots F_{k_{Bi}}$ in the Cartesian space $F_1 \times F_2 \times_{\ldots\ldots} F_k$. Finding the cost of energy usage involves finding the one to one majority function ***f:*** $fF_1 \times F_2 \times \ldots\ldots F_x \ldots\ldots R$. The mapping produced by this function ***f*** i.e. the cost of energy usage in different sport centres, is determined from past experience but a precise functional form or a mathematical model of this is not known. This is an approximation to this feature space, which can be defined. There are three possible methods for modelling the function ***f***. In this work, regression models *(y=f(x))*, neural networks and neuro-fuzzy are used.

DATA USED IN THE EXPERIMENT

In order to develop a cost model, a sample of data was collected from Liverpool City Council: Directorate of Leisure Services. The data was collected by documentary sources as opposed to others highlighted by Robson (1993) such as direct observation, mailed questionnaire and interviewing.

Fellows and Liu (1997) describe the fundamentals associated with data collection and in particular highlight the need to investigate the nature of data and collection mechanisms in order to be aware of the limitations of the data and their validity, notably comparability. As mentioned previously, data was extracted directly from a database provided by Liverpool Council. This method of data collection has the advantage of “value-free” research. That is to say that the work is unaffected by the beliefs or intentions of the researchers or data sources. Similarly, as the research is not based upon the data collected from previous work, then the actual data will be raw and will not have been tailored for other purposes.

SELECTING THE VARIABLES

Based upon statistical pre-analysis, variables were selected for predicting the cost of energy usage. The influence of variables on the cost of energy and ease of availability of data were the primary factors in the selection of the variables. The experiment uses floor area and number of user variables to predict the cost of energy usage. The correlations of the chosen data were all small and hence the chosen variables are independent.

REGRESSION MODELS

The relationship between energy cost and sport centres characteristics can be modelled in the general input / output function;

$y=f(x)$ *eqn.(1)*

then applying the selected variables to the function;

y(*total*cost) = f(*floorarea, user*) or

$y = f(x_1x_2)$………*eqn.* (2)

Given the floor area and number of users over a period of time, in this case one year, it is possible to predict the total cost of energy over the same period. Many techniques can be employed to estimate the parameters or coefficients of the relatives in Equation 1. In this work, two methods were employed. First it was assumed that the relationship in Equation 1 is of a linear type if the following form;

$$y = c + b_1 x_1 + b_2 x_2 \,..........\, eqn.(3)$$

Then multiple linear regression (an expansion of the least square regression method in that it involves more than one independent variable), is used to determine the coefficients. Second, a non-linear form for Equation 2 was investigated, because some of the variables might have a non-linear relationship with the total cost of energy. In total, five non-linear combinations were developed and tested. The equations of these models are as follows:

$$y = c + e^{b_1 \log x_1} + e^{b_2 \log x_2} \,........................\, eqn.(4)$$

$$y = c + e^{b_1 \log x_2} + e^{b_2 \log x_1} \,........................\, eqn.(5)$$

$$y = c + e^{b_1 \log x_1} + b_2 \log x_2 \,.....................\, eqn.(6)$$

$$y = c + b_1 (\log x_2)^{a_1} + b_2 \log x_1 \,............\, eqn.(7)$$

$$y = c + b_1 (\log x_2)^{a_1} + b_2 (\log x_1)^{a_2} \,............\, eqn.(8)$$

SPSS Statistical Package was used for multiple linear regression analysis to form a set of regression coefficients and their respective ***t***- statistics. The ***t***- statistics were significant for every regression coefficient. The regression coefficients obtained were used to predict the cost of energy usage of both the learning sample (to see how well the regression model fitted the learning sample) and the testing sample (to see how well the regression coefficients generalize and predict the cost of energy usage). The results of this testing showed that the linear model fits data very well. Therefore this model was selected for comparison against ANN and neuro-fuzzy models. Other statistical measures are used to measure the predictability of the developed models. The results will be discussed in the next section.

NEURAL NETWORK MODELS

ANN modelling differs from statistical modelling in the sense that in regression models the approximated function is assumed and the regression coefficients are calculated, where as ANN is a function itself is asked to approximate the unknown function that maps the input space to output. In other words, a function is being asked to approximate another function. The model is said to be able to solve the problem if it is to learn to approximate the function to an arbitrary accuracy. (For the sake of clarity, this is an abbreviated description of ANN. For more detailed information, refer to the references).

To find this approximated function several experiments, with different ANN configurations were developed. The RMS error was kept constant for different neural network configurations. The performances of the various neural network configurations were tested against the learning sample and testing data set. Various error measurement techniques were used for selecting the optimum predictive model. The results are explained and summarized in the next section.

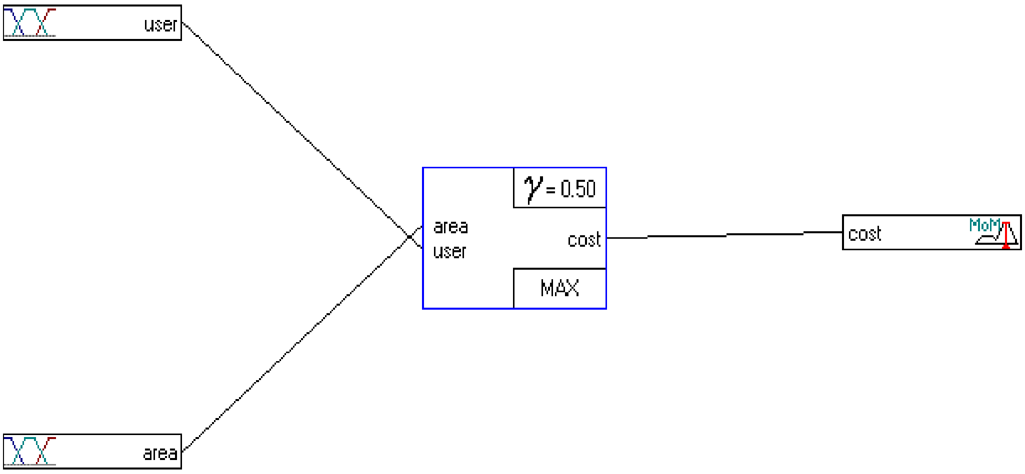

Figure 1 Structure of Neuro-fuzzy system

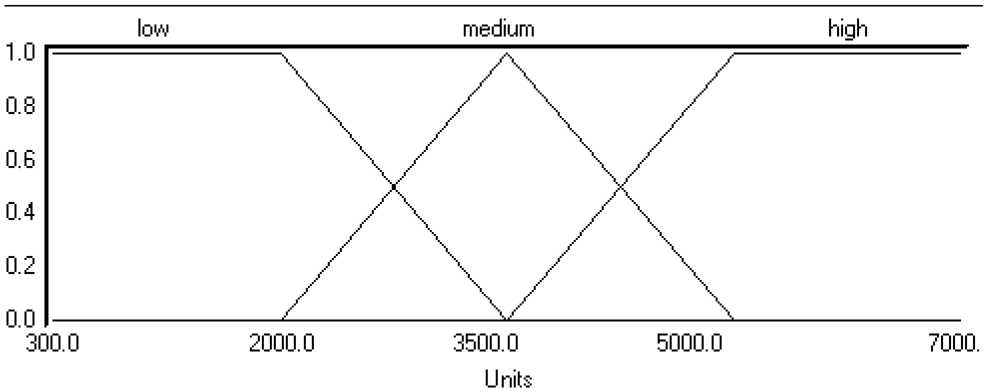

Figure 2 Membership function of input variable: area

NEURO-FUZZY MODELLING

Finally, a forecasting model was prepared using the fuzzy logic. Neuro-fuzzy is a combination of the explicit knowledge representation of fuzzy logic with the learning power of neural networks. Neuro-fuzzy modelling involves the extraction of rules from a typical data set and the training of these rules to identify the strength of any pattern within the data set. The system creates membership functions from which linguistic rules can be derived as opposed to real values.

Figure 1 shows the fuzzy system used to create the membership functions for the input and output variables shown in Figures 2, 3 and 4. For this purpose of cost modelling, many alternative methods of integrating neural networks and fuzzy logic have been proposed in literature. Amongst this method is Fuzzy Associate Memories (FAM). FAM is a fuzzy logic rule with an associated weight. This method is based on a mathematical function that maps FAMs to neurons in the neural network. This enables the use of a modified error back propagation algorithm with fuzzy logic.

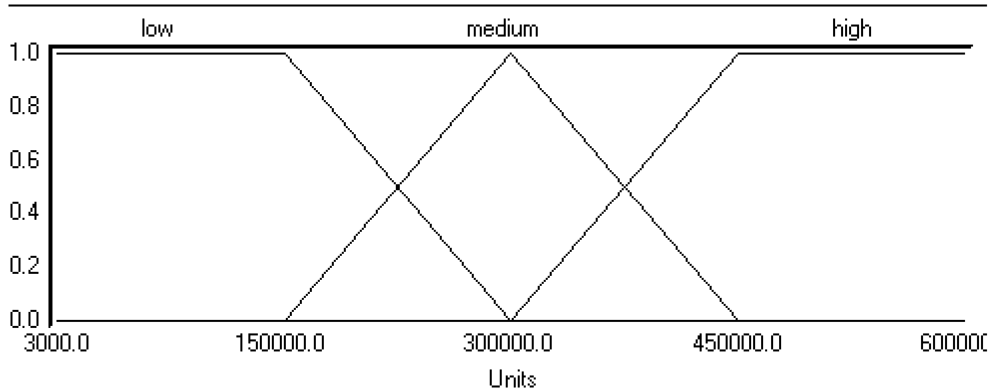

Figure 3 Membership function of input variable: users

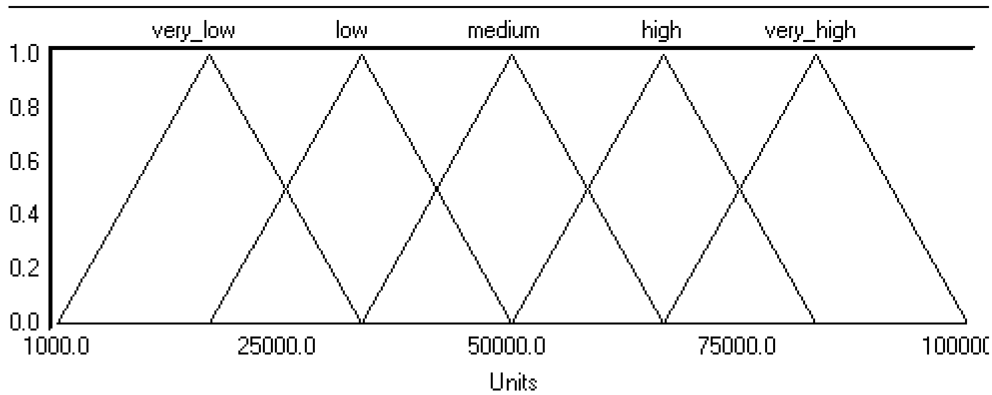

Figure 4 Membership function of output variable: energy cost

TESTING METHODOLOGY

The purpose of the comparison is to evaluate the performance of three modelling techniques in estimating a functional form that relates the characteristics of sport centres to the cost of energy usage. To perform this kind of comparison social scientists advocated testing alternative models side-by-side in critical experiments. Therefore, this work compares side-by-side the performance of the three modelling techniques. Since the true underlying functional form is unknown the performance of three selected models was examined. The performance of the three models was evaluated across each other. The mean absolute percentage error (MAPE) and Theil's U statistic was used to measure the performance. To assert the finding of these measures the paired t-test was used to test the difference between the three models. Taken into account the assertion that neural networks were able to see through noise and distortion and extract the functional form automatically, it was expected that the ANN model would out-perform the other models.

Paired Difference t-test

The easiest way to test for differences between the three models is to calculate the difference between each two pair of models. This will result in three sets of observations. These are:

1. ANN Vs Regression
2. ANN Vs Neuro-fuzzy
3. Regression Vs Neuro-fuzzy

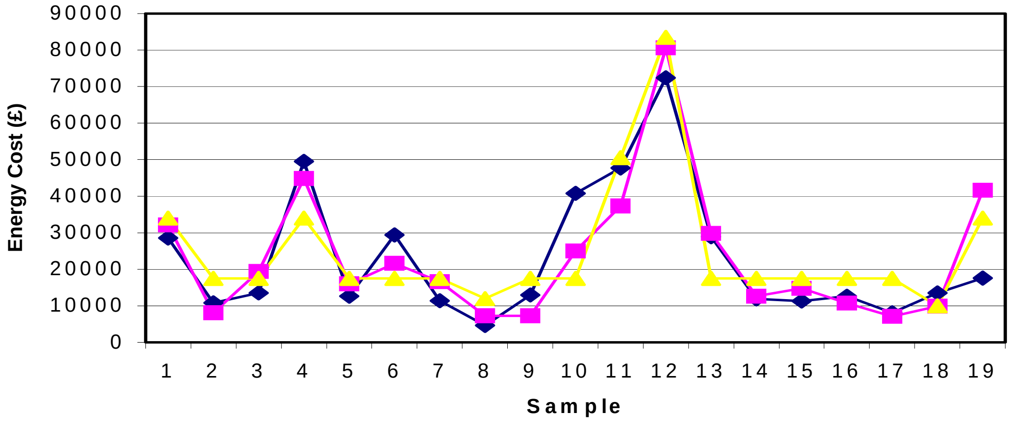

Figure 5 Comparison of MAPE

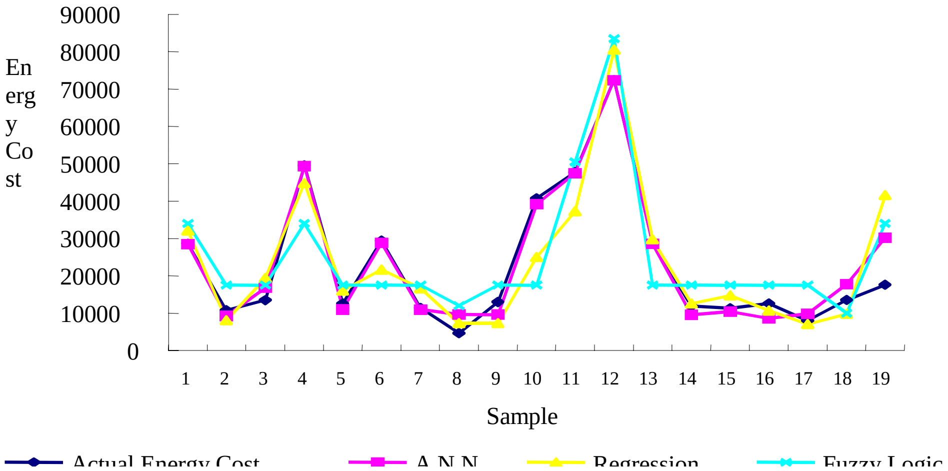

Figure 6 Actual v forecasted cost

If any of the above two sets of models are similar, then the mean difference is expected to be around zero. But if the mean difference is much bigger than zero, then there will be a real difference between the above sets of models. This require to test the null hypothesis that the MAPE mean difference is zero. The following hypothesises and alternative hypothesises were tested:

1. H_0: $\mu_1 = 0$ the difference between ANN and Regression models MAPE is small
2. H_1: $\mu_1 \neq 0$ the difference between ANN and Regression models MAPE is large to suggest a real difference
3. H_0: $\mu_2 = 0$ the difference between ANN and Neuro-fuzzy models MAPE is small
4. H_2: $\mu_2 \neq 0$ the difference between ANN and Neuro-fuzzy models MAPE is large to suggest a real difference
5. H_0: $\mu_3 = 0$ the difference between Regression and Neuro-fuzzy models MAPE is small
6. H_3: $\mu_3 \neq 0$ the difference between Regression and Neuro-fuzzy models MAPE is large to suggest a real difference

Table 1: Results of paired two sample for means t-test

t-Test: Paired Two Sample for Means	**ANN vs Regression**		**ANN vs Neuro-fuzzy**		**Regression vs Neuro-fuzzy**	
	MAPE ANN	MAPE Reg.	MAPE ANN	MAPE Fuzzy	MAPE Reg.	MAPE Fuzzy
Mean	20.03346	31.21781	20.03346	50.61951	31.217808	50.61950548
Variance	760.4342	874.7848	760.4342	1357.131	874.78476	1357.131074
mean diff.	11.1844		-30.586		-19.4017	
Std. dev. of diff	24.759		23.74		36.052	
SE of Mean	5.446		5.446		8.271	
Pearson Correlation	0.626657		0.764833		0.4277727	
Hypothesized Mean Difference	0		0		0	
df	18		18		18	
t Stat	-1.96904		-5.61579		-2.345799	
P(T<=t) one-tail	0.03227		1.25E-05		0.0153209	
t Critical one-tail	1.734063		1.734063		1.7340631	
P(T<=t) two-tail	0.06454		2.5E-05		0.0306417	
t Critical two-tail	2.100924		2.100924		2.1009237	

The results of this trail are presented in Table 1 and will be discussed in the results.

RESULTS

The regression models shown earlier in the paper were tested for accuracy by calculation of predicted results and Mean Absolute Percentage error. A neural network model was also produced which recorded a correlation of 0.9579 (RMSE 0.1167), this accuracy being better than that recorded by the regression models. Finally, a Neuro-fuzzy model was also produced through the rule block shown in Fig 1. Although the Neuro-fuzzy model recorded the lowest accuracy level, the model was able to extrapolate better than the two other models as demonstrated by sample 18 from the testing set. Figure 5 compares the MAPE results of the three models. The figure shows that the MAPEs of ANN model range between 0-20 except for sample 8 where the MAPE was 150. The MAPEs of the regression model ranged from 0-71 bar sample 8 whereas the neuro-fuzzy model MAPEs range from 0-158. The average MAPEs are 20%, 31% and 50% for ANN, regression and neuro-fuzzy respectively. The Theil's U value for the regression model is 0.142 , for the ANN model is 0.0685 and for the neuro-fuzzy model is 0.348 (where the U value indicates greater accuracy as $U \rightarrow 0$).

Table 1 summarizes the results of paired difference t-test. The results indicate that the ANN model and regression model are very close whereas there is no similarity between ANN and fuzzy models. The difference in MAPE is 10% between ANN and regression, 30% between ANN and neuro-fuzzy and 20% between regression and neuro-fuzzy models.

DISCUSSION

The overall MAPEs for ANN model was acceptable. Also the mean difference between regression and ANN was around 10 %. Even when there were statistically significant difference between ANN and neuro-fuzzy, and regression and neuro-fuzzy models, these differences may not hold any practical significance. The level of MAPE for ANN and regression model can be considered acceptable in most real applications, depending on the phase of application of the model. The t-test indicated that the ANN estimates did not differ significant from ANN whereas there were a significant difference with the other models. It is clear from this limited experiment that ANN were able to extract a functional form (i.e. a function) that represent the problem under investigation. This study considered only few measures. It would be interesting to examine the impact of outlying or highly influential data on these modelling methods the other models

CONCLUSION

The paper has highlighted the importance of different modelling techniques for predicting energy consumption levels in new and existing sport centres. The study has also shown that ANN models may prove as a good alternative to regression and neuro-fuzzy as far as functional form extraction is concerned. Within the limits of this study, ANN models have been shown to be able to model data that strongly exhibit noise and achieve reasonable accuracy.

REFERENCES

Fellows, R and Liu, A. (1997) *Research methods for construction*. Oxford: Blackwell.

Robson, C. (1993) *Real world research*. Oxford: Blackwell.

Kirkham *et al.* (1999) Forecasting the running costs of sport and leisure centres. To be published *In: Procs 8th annual conference of durability of building materials*, Vancouver, Canada. National Research Council Canada.