

A neural network-based methodology of quantifying the association between the design variables and the users' performances

Abstract—User performance is highly correlated with design variables of a system. Such association can be described as display-control relationship. In this study, a neural network-based methodology is proposed to identify and quantify the association among design variables (inputs), and to compute their relative influences on the two performance measures (outputs) of user response time and response accuracy, using artificial neural network (ANN), **generalized regression neural network (GRNN)**, support vector regression (SVR), multiple linear regression (MLR), and response surface model (RSM). Based on the results of the comparison, it is found that neural network-based methods are more reliable than support vector regression-based methods in computing the relative influence of design variables. As a result of our analysis, the best option for optimizing each of the measures is suggested. Some useful observations about the design of man-machine systems are also presented, discussed and visualized. In the study of man-machine systems, quantitative methods are seldom adopted for examining the mappings between various displays and controls under a variety of operating conditions. The major contribution of this study is to provide some insights into the usefulness of quantitative methods in evaluating man-machine design in terms of display-control compatibility, and to extract explanatory information from renowned black-box systems such as neural networks.

Keywords: neural networks; display-control compatibility; support vector regression; modelling.

1. INTRODUCTION

When operating man-machine systems, two-way communication between the users and machines must be facilitated via the proper use of displays and controls. The relationship between a control motion and its causal effect expected by a population can be defined as a direction-of-motion stereotype. Such a stereotype is said to be compatible if it can match the expectation of the majority (Chan and Chan, 2008a). Since advancement of mechanical and electronic mechanisms has evolved over the past few years, it further complicates the connection between displays and controls. For instance, a rotary knob can be used to command the translatory movement of an indicator. Likewise, a two-way lever can control rotary movement on a display. As a result, a vast number of possible display-control relationships can be observed. In many sophisticated systems, not all display-control relationships can be explicitly and consistently defined. Even worse, two opposite types of control motions are sometimes required for the same type of display effects in some industrial practices (Chan and Chan, 2008b). Since incompatible display-control mappings can result in serious mishaps and accidents (Burgess-Limerick et al., 2010; Burgess-Limerick et al., 2012), it is useful to identify how such relationships may affect the users, and also to make suggestions to predict human activities (Kang and Seong, 2001). In this connection, the study of display-control relationships under different operating conditions has been attracting a lot of researchers over the past few decades (Chua et al., 2001; Courtney, 1992, 1994a). For example, some studies have been dedicated to specific display-control relationships, such as Translatory-Linear (Worringham and Beringer, 1998), Rotary-Linear (Brebner and Sandow, 1976; Courtney, 1994b; Hoffman, 1997), Rotary-Circular (Chan and Chan, 2006; Chan and Chan, 2003), Translatory-Circular (Chan and Chan, 2007) and even Translatory-Digital (Chan and Chan, 2008a). However, the participants in all the above studies were always instructed to manipulate the controls on the frontal plane. A study on the operation of overhead travelling cranes has found that a quarter of the controls were not placed in front of, but on the left and right sides of the users. These incompatible control arrangements could have induced more human errors and serious accidents, especially in an emergency condition or under a high workload (Sen and Das, 2000). For some workplaces, like cockpits and nuclear control rooms, that contain a large number of controls and displays, it may be difficult to locate the optimal display-control placements such that users can have comfortable access, easy information detection and straightforward response execution (ISO, 1998). Such variations in relative orientations of users, controls, and displays of interest are also seen

in automobiles (Makiguchi et al., 2003), unmanned air vehicles (McCarley and Wickens, 2005), and coal mine shuttle cars (Zupanc et al., 2007). Due to the fact that there are very few studies dedicated to a working environment with controls in front of the users and on multiple displays of interest, this paper intends to fill the gap. Under such environments, the users may need to first transform the display-control spatial mapping into their expected stereotype and then operate a control device for an expected effect. According to our knowledge, the principle of such mapping has not yet been fully explored.

In the study of display-control compatibility, the users may be required to exercise translatory or circular motion in operating a control device rather than simply pushing or pulling a control. Direction-of-motion stereotypes can be investigated with a variety of control devices (e.g. two-way lever, rotary knob) and common displays (e.g. horizontal scale, vertical scale, circular scale). Differing from the previous studies on stereotype compatibility, this study will mainly examine the connectivity of different display-control relationships and the reactions of the participants under a setting that a subject would react to a given instruction by operating a device on a fixed control plane to observe an expected effect on a display oriented in different directions. In this study, three control devices are provided for users to execute six instructions on three display scales located at the four cardinal orientations. A number of experiments are then performed to measure the performance of the users under different operating conditions. Therefore, the primary objective of this study is to find out which design variables are more influential and how they affect the response of the participants using quantitative methods. By doing so, it is believed that some useful design guidelines can be identified to enhance the user performance.

Standard correlation analysis is used to determine which variables are significant. Some sophisticated techniques are then employed to model the association between the significant design variables and the performance measures. Among various quantitative modeling approaches, because of the ability of artificial neural network (ANN) to provide subjective explanations and evaluations, it has been utilized to support decision-making and problem-solving in a wide range of domains, including stock market forecasting (Cao et al., 2005; Enke and Thawornwong, 2005; Kim and Shin, 2007), credit rating (Huang et al., 2004; Yu et al., 2008), ecological science (Olden and Jackson, 2002; Olden et al., 2004), consumer analysis (Deng and Pei, 2009; Deng et al., 2008; Hu and Tsoukalas, 2003), data mining (Kewley et al., 2000) and human movement modeling (Kawato et al., 1990; Rigotti et al., 2001; Uno et al., 1989; Wong and Chan, 2012). By making use of the mathematical utility of ANN, we can generate predictive (or regression) models of a domain problem and then develop some explanatory methods to quantify the influence of the model variables. One way is to extract the input influence from the connection strengths (inter-layer weights) of a trained ANN model using Garson's method (Garson, 1991), Yoon's method (Yoon et al., 1994) and Tsaur's method (Tsaur et al., 2002). However, some of these methods may subject to its own deficiencies (Howes and Crook, 1999) and they may produce different interpretation results given the same connection strengths like the one reported in Wong et al. (2011). Alternatively, one may change the model inputs and observe the associated effects on the model outputs so as to statistically determine the input influence using change of mean square error (COM) and sensitivity analysis (SA). The major advantage of the above two methods is that they can be easily applied to modeling techniques other than ANN. For ease of presentation, the details of COM and SA will be explained in Section 4.5.

Recently, a new regression method known as support vector regression (SVR), has been studied and applied to a variety of domains such as systems reliability forecasting (Chen, 2007), order flow-time prediction (Alenezi et al., 2008), tourism demand forecasting (Chen and Wang, 2007), human performance modeling (Bi et al., 2011), and human perception modeling (Zhu et al., 2007). Unlike conventional regression methods (e.g. multiple linear regression) and ANN, SVR models can be developed and trained by minimizing the upper bound of the generalization error rather than the training error. In addition, SVR does not allow the searching process to be trapped at local minima. Like ANN, SVR has been also applied to compute the input influence (e.g. Cortez et al., 2009; Cortez et al., 2006). Despite the fact that ANN has been commonly used in the computation of the input influence, there is still no consensus about which method is the best. Hence, we decide to examine other modeling techniques, **generalized regression neural network (GRNN)** and SVR and attempt to compare them with ANN in this study. According to our knowledge, there are also very few studies dedicated to the comparison between ANN, GRNN, and SVR in the computation of the input influence. Specifically, as the primary objective, ANN, GRNN, SVR and multiple linear regression (MLR) methods are adopted to quantify the causal connection between the design variables and the user performance measures, and to compute the relative influence of variables. Some useful suggestions are then presented and discussed based on the comparison results. In addition, since the application of nonlinear methods (i.e. ANN,

GRNN, and SVR) has not been fully explored in the domain of ergonomics, the secondary objective of this study is to provide some insights on the usefulness of nonlinear methods in modeling display-control compatibility.

This paper is organized as follows. The research background is presented in Section 2. Section 3 describes the use of statistical methods to develop relation graphs among variables. Section 4 reports the use of different regression methods to quantify the causal relationship between variables, and the computation of their relative influences. Section 5 discusses the usefulness of the results obtained. The paper is concluded with the future research direction in Section 6.

2. RESEARCH BACKGROUND

2.1 Objective

An attempt has been made to investigate the connectivity of different display-control relationships and the performance of the users. Unlike our previous study (Wong and Chan, 2012) in which each participant was asked to interact with one rotary control device installed on three different planes and to observe an effect oriented in three different directions, we decide to pay more attention on the mapping between control devices and instructions. Therefore, as an extension of our previous study, each participant is now requested to use three types of control devices to execute six different instructions that appear on displays oriented in four different directions. The performance of the subject is then recorded and analyzed. For ease of presentation, all variables in this study can be divided into two categories, controllable inputs and measured outputs, as shown in Table 1. Controllable inputs refer to the factors that can be controlled, adjusted, or manipulated intentionally. These inputs include the type of control device used (CD), the orientation of the display (DO), and the instruction that the participants need for execution (IN). Measurable outputs, on the other hand, refer to the reaction of the participants after they are exposed to the inputs. These outputs include the actual response of the participants regarding the given instruction (RP), the reaction time of the participants (RT), and the accuracy of the actual response (RA).

“Table 1”

2.2 Experimental Setup

To achieve the research objective, a number of experiments were performed with different control devices and display orientations. The participants were asked to use their right hand for manipulating the control device at a fixed location (in front of the North display) as shown in Figure 1. There were three types of control device (CD), i.e. horizontal, vertical, and rotary. A 2-way lever joystick was regarded as both a horizontal and vertical control device while a 35 mm knob was used as rotary control device. The displays were set at a distance of 750 mm away from and at 15° degree below the eye level of the participants. There were four display orientations (DO) relative to the users, namely North (N), East (E), South (S), and West (W). The dotted lines represent the line of sight of the users. The tests were performed on 12 blocks of three control devices and four display orientations. In this connection, the participants could observe the target display in different orientations by manipulating different control devices using their right hand.

For each block of tests, three display scales were used, as shown in Figure 2. At the beginning of the test, a pointer was always placed at the mid-point of the horizontal (vertical) scale or at the 12 o'clock position on the rotary scale. Using one of the control devices, the users were requested to complete six different instructions (IN): (1) Moving the pointer to left (L); (2) moving the pointer to right (R); (3) lifting the pointer up (U); (4) putting the pointer down (D); (5) turning the pointer in the anti-clockwise (AC) direction; and (6) turning the pointer in the clockwise (C) direction. In normal practice, instructions (1) and (2) should be performed on a horizontal scale (Figure 2a). Likewise, instructions (3) and (4) must be executed on a vertical scale (Figure 2b), while the rotary scale is for instructions (5) and (6) (Figure 2c). However, these display-control relationships may not hold all the time. For example, it is quite common to use a single knob to turn one pointer in the AC/C direction and move another pointer to L/R horizontally. In this study, different display-control relationships are examined where the participants' performance can be analyzed. To better investigate the performance of the participants under study, the response (RP) of the participants must be recorded, and the response time (RT) and response accuracy (RA) of the participants can then be captured. RT, as measured in millisecond (ms), was counted as the difference between the instruction time and the action time of the

participants. RA was measured in two levels: (1) False if the control motion of the users is opposite to the requested direction; and (2) otherwise, true. If horizontal (vertical or rotary) control is used for a horizontal (vertical or rotary) scale, such display-control relationship is said to be highly compatible and RA can be easily measured such that $RA = \text{False}$ if $IN \neq RP$, otherwise $RA = \text{True}$. However, when a lever joystick is used for a rotary scale or a rotary knob is used for a horizontal/vertical scale, RA must be measured with respect to the direction-of-motion stereotypes. The details are given in Section 3.1. In all tests, the pointer always moved in the requested direction regardless of the direction of the control motion of the users.

In summary, there were totally 72 (12 blocks x 6 instructions) test conditions x 4 repetitions = 288 trials. Thirty-seven university students (18 male and 19 female) aged between 20-24 years took part in this experiment. They are all right-handed and were screened by the four-item Lateral Preference Inventory of Coren (1993). Hence, the total number of dataset required is 10656 (288 trials x 37 participants). However, only 10486 datasets were obtained due to some technical problems such as unexpected errors during data transfer. Nevertheless, our analysis was done based on the error-free datasets.

“Figures 1-2”

3. DEVELOPMENT OF RELATION GRAPHS

Correlation analysis and partial correlation analysis were applied to describe the causal association among variables in the experimental data. The analysis results are used to construct the relation graphs for two major outputs, response time (RT) and response accuracy (RA). Relation graphs can explicitly reveal the connection between the inputs and the two outputs. Based on the relation graphs, the relative influence of inputs toward outputs can then be computed. As a result, it is possible to know which input is the most influential and how such an input can be mastered to achieve a desirable output with a higher confidence level.

3.1 Correlation Analysis

Using the datasets collected from the experiment, the one-to-one relationship between all factors can be statistically investigated via correlation analysis. Generally speaking, it is desirable to observe significant correlations between inputs and outputs. Among the two major outputs, RT can be recorded easily. In contrast, RA cannot be directly measured from the experimental data when the display-control relationship is not highly compatible (e.g. horizontal-vertical, vertical-rotary). Hence, in such circumstances, RA can be only measured with respect to the direction-of-motion stereotypes that the majority would follow. The actual responses of all users under different display-control relationships and display orientations are reported in Tables 2-4. From Table 2, the stereotypes of horizontal control can be expressed in form of IN-RP: L-L, R-R, U-R, D-L, AC-L, and C-R. It means that if the L (or R) instruction is given, the users tend to move the lever to L (or R). Also, if the U (or D) instruction is requested, the majority tend to move the lever to R (or L). Likewise, the users tend to move the lever to L (or R) for AC (or C) instruction. The stereotypes of vertical and rotary controls can be observed from Tables 3 and 4 respectively. In short, the stereotypes of vertical control are: L-D, R-U, U-U, D-D, AC-D, and C-U. For rotary control, the stereotypes are: L-AC, R-C, U-C, D-AC, AC-AC, and C-C. Therefore, for incompatible display-control relationships, if the control motion of the users is the same as the stereotype of that particular control device, $RA = \text{True}$, otherwise, $RA = \text{False}$.

After we applied correlation analysis to the experimental data, the correlation coefficients (r) are listed in Table 5. It is clear that there is no significant correlation between all controllable inputs which means they are independent variables. Among the two major outputs, RT negatively ($r = -0.0466$, $n = 10486$) and significantly ($p = 0$) correlates with RA. This indicates that a longer reaction time may be associated with a higher rate of having the wrong reaction to a given instruction, and vice versa. Next, the connectivity between the inputs and outputs must be examined. If such connection exists, these inputs can be used to control or predict the outputs. From Table 5, weak but significant correlations ($p = 0$) among CD-RT ($r = 0.0679$, $n = 10486$) and CD-RA ($r = 0.055$, $n = 10486$) can be observed. Significant ($p < 0.01$) but weak correlation ($r = 0.0289$, $n = 10486$) can be noted between DO and RA. Moreover, significant correlation ($p = 0$) can be addressed between IN and RA while a marginally significant association ($p < 0.05$) exists between IN and RT.

“Tables 2-5”

3.1.1 General Observations

Among all remarkable associations, RT positively correlates with CD and IN. The link CD-RT suggests that the change of control device from horizontally moving to rotary could impose additional mental effort while IN-RT may merely indicate the normal reaction time of the participants. RA positively correlates with CD and DO but negatively correlates with IN and RT. CD-RA implies that most of the participants can react correctly to a given instruction using different control devices. DO-RA means that the display orientation has little impact on the response accuracy. IN-RA may indicate that the users have difficulty in executing U/D/AC/C instructions using different control devices. The link RT-RA may also imply that if participants need to exercise extra mental effort, longer reaction times and incorrect responses are more likely to be observed. Accordingly, an initial relation graph depicting important input-output correlations can be constructed as shown in Figure 3.

“Figure 3”

In Figure 3, a solid line indicates that there is a remarkable correlation between two variables whereas a dotted line is a presumed partial correlation between two variables. Partial correlation is generally used to depict the relationship between two variables while controlling for a third variable. It aims at addressing the unique variance between two factors while eliminating the variance from the third one. If the partial correlation is significant between two factors for controlling the third one, the association between the two factors is unique and essential. Otherwise, the association between the two variables is greatly affected by the third one. If this is the case, no direct linkage (solid line) is established between them.

3.2 Partial Correlation Analysis

Using partial correlation analysis, each connection between two variables (in Figure 3) was re-verified for controlling the third variable. If such connection was still found to be remarkable, the association would remain, otherwise, the linkage should be omitted (Wong et al., 2011). As a result of analysis, all associations were deemed significant or marginally significant (at least $p < 0.05$) except the link between IN and RT for controlling RA. It implies that the connection between IN and RT was greatly affected by RA. Hence, that linkage must be removed. Based on the results of partial correlation analysis, the relation graph can be revised as shown in Figure 4. From the revised graph, two important hubs regarding two major outputs, RT and RA, can be located. Since there is a mutual connection between the outputs, RT should be treated as a contributing input to RA and vice versa. Hence, two relation graphs are developed, one for RT with two inputs while another for RA four inputs. Next, regression models can be developed to measure these input-output connections.

“Figure 4”

4. DEVELOPMENT OF REGRESSION MODELS

4.1 Artificial Neural Network (ANN)

ANN is used to model the relationship between inputs and outputs. For a typical ANN, each line in the structure of a fully trained ANN model is endowed with a so-called connection weight specifying the connection strength from one unit to another (Bishop, 1995). A typical three-layer ANN can be viewed as a non-linear regression model with an identity transfer function in the output unit and logistics transfer functions in the middle-layer units. Such a model can approximate any continuous function arbitrarily well, given that a sufficient number of units appear in the middle layers (White, 1990; Yu et al., 2008). If readers would like to know more details about ANN, please refer to the text of Bishop (1995) and Lek et al. (1996). For benchmarking purposes, generalized regression neural network (GRNN), support vector regression (SVR), multiple linear regression (MLR), and response surface model (RSM) were applied. Based on the two relation graphs, two models can be developed. For the first model (M1), there are two inputs (CD and RA) and one output (RT). There are four inputs (CD, DO, IN, and RT) and one output (RA) for the second model (M2).

Using the experimental data, ten-fold cross-validation was performed to train and test the ANN for both M1 and M2. By doing so, a robust estimation of the quality of the inputs would be obtained (Cortez et al., 2006). As suggested by Huang et al. (2004), for each of the ten cross-validation processes, 90% of the datasets ($10486 \times 0.9 = 9438$) were used to train the ANN and the remaining 10% ($D = 1048$) were used to test the ANN. The same procedure of training and testing was also applied to GRNN, SVR, MLR and RSM. Using ANN, the number of neurons (P) in the hidden layer must be determined first. A trial-and-error method was performed to optimize the value of P (Wang and Elhag, 2007). The performance of the ANN for M1 was measured by the average mean absolute percentage error (MAPE) and standard deviation (SD) of the average MAPE over ten validations. MAPE is defined by Equation (1) where A_i and E_i are the actual values and the predicted values of the i -th test dataset respectively, and D is the total number of test datasets. The performance of the ANN for M2 was measured by the prediction correctness (PC). Given that RA is a binary output (0 or 1), PC is defined as the ratio of the total number ($\sum B$) of correct predictions (when $A_i = E_i$) over the total number of test datasets (D), as illustrated in Equation (2). Table 6 shows the average performance of the ANNs with varying P for M1 and M2, over ten cross-validation processes. In general, it can be observed that ANN is improved as P increases. Although a large P may increase the ability of ANN to learn and memorize the data, an over-fitting problem may be induced. For M1, the ANN with the smallest MAPE and SD, is considered to be the best in predicting RT. Hence, the ANN with P = 6 is selected for M1. For M2, the ANN with P = 4 yields the highest PC in predicting RA.

$$MAPE = \frac{1}{D} \sum_{i=1}^D \frac{|A_i - E_i|}{A_i} \times 100 \quad (1)$$

$$PC = \frac{1}{D} \left(\sum_{i=1}^D B_i \right) \times 100 \quad (2)$$

$$\text{where } B_i = \begin{cases} 0 & \text{if } A_i \neq E_i \\ 1 & \text{if } A_i = E_i \end{cases}$$

“Table 6”

4.2 Generalized Regression Neural Network (GRNN)

Unlike ANN, GRNN, developed by Specht (1991), is free from extensive experimentation and investigation to develop the network structure. A typical GRNN consists of four layers namely, the input layer, the radial basis (pattern) layer, the summation layer, and the output layer. Units in the input layer are fully connected to the radial basis layer, whose output is a measure of the distance of the input from the stored pattern. Each unit in the pattern layer is connected to two units (S and D summation neurons) in the summation layer. The S neuron measures the sum of weighted outputs of the pattern layer while the D neuron computes the unweighted outputs of the units in the pattern layer (Kim et al., 2010; Haidar et al., 2011). Using GRNN, the spread (smoothing factor, σ) which can affect the degree of generalization of the network must be determined first. If σ is large, the network can gain ability to generalize but the prediction may be degraded. In this study, σ is determined using a trial-and-error method. Table 7 reports the average performance of GRNN with varying σ for M1 and M2, over ten cross-validation processes. In this connection, GRNNs with $\sigma = 1.4$ are considered to be the best network structure for both M1 and M2.

“Table 7”

4.3 Support Vector Regression (SVR)

A relatively new modeling method, SVR, is an application of the support vector machine (SVM) which is a group of supervised learning techniques that can be used for both classification and regression. Using SVR, the inputs are first nonlinearly mapped into a high dimensional feature space (F), wherein the inputs are linearly associated with the outputs (Lu et al., 2009). The SVR formulation can be expressed in Equation (3) where w is the weight vector, $\Phi(x)$ is a mapping function, and b is a constant. Among various loss functions, the most popular one is the ϵ -insensitive loss function (Vapnik, 2000), as defined in Equation (4), where ϵ is used to specify the region with less than ϵ deviation

from the actual output values for all the training data. Hence, the loss function does not penalize the prediction errors within this “ ε -insensitive” region. One major advantage of using this loss function is that it makes practical computation feasible by producing a sparse set of support vectors. For more details about loss functions, please refer to the study of Smola and Scholkopf (2004). According to Vapnik (2000), the weight vector (w) and constant (b) can be predicted by minimizing the regularized risk function as shown in Equation (5) where $\|w\|^2/2$ is the regularization term, C is the regularization constant used to determine the tradeoff between the regularization term and empirical risk, and D is the number of training samples. This minimization function can be transformed into a dual problem, and the general form of the regression function can be expressed in Equation (6), where α_i and α_i^* are Lagrange multipliers for the i -th training sample, and $K(x, x_i)$ is a kernel function satisfying Mercer’s conditions (Vapnik, 1999, 2000). The value of the kernel equals the inner product of two vectors, x_i and x_j , in the feature space $\Phi(x_i)$ and $\Phi(x_j)$, i.e. $K(x, x_i) = \Phi(x_i)\Phi(x_j)$. Although several kinds of kernel functions have been specified and made available, the most widely adopted is the radial basis function (RBF), as defined in Equation (7), where λ represents the width of the RBF.

$$f(x) = w \cdot \phi(x) + b \quad (3)$$

$$L_\varepsilon(f(x), y) = \begin{cases} |f(x) - y| - \varepsilon & \text{if } |f(x) - y| \geq \varepsilon \\ 0 & \text{Otherwise} \end{cases} \quad (4)$$

$$R(C) = \frac{1}{2} \|w\|^2 + \frac{C}{D} \sum_{i=1}^D L_\varepsilon(f(x_i), y_i) \quad (5)$$

$$f(x) = \sum_{i=1}^D (\alpha_i - \alpha_i^*) K(x, x_i) + b$$

Subject to

$$\sum_{i=1}^D (\alpha_i^* - \alpha_i) = 0 \quad (6)$$

$$0 \leq \alpha_i, \alpha_i^* \leq C$$

$$K(x, x_i) = \exp\left(\frac{-\|x - x_i\|^2}{2\lambda^2}\right) \quad (7)$$

It is well-known that the modeling accuracy of the SVR depends heavily on the hyper-parameters C , ε and the kernel parameters. If RBF is adopted, the width (λ) of the RBF must be determined. In this study, both C and ε are addressed using the heuristics proposed by Cherkassky and Ma (2004). The heuristic used to determine C is described in Equation (8) where \bar{y} and σ_y are the mean and standard deviations of the output values (y) of the training data. Equation (9) is used to set ε , where σ is the standard deviation of the noise and D is the number of training samples. According to Cherkassky and Ma (2004), the noise variance (σ^2) can be readily estimated by fitting the data using a low bias model (e.g. MLR, high-order polynomial) and applying formula (10) to determine σ^2 (Cherkassky et al., 1999). In formula (10), d is the degree of freedom of the estimator and D is the number of training samples. Using Equations (8)-(10), $C=1.34$ and $\varepsilon=0.02$ can be computed for M1 while $C=2.87$ and $\varepsilon=0.03$ for M2. Also, four different values of the RBF width (λ) are compared for both M1 and M2, as reported in Table 8. Thus, $\lambda=0.1$ is used for M1 given that MAPE and SD are the smallest while PC is quite insensitive to λ for M2, so λ is also set to 0.1.

$$C = \max(|\bar{y} + 3\sigma_y|, |\bar{y} - 3\sigma_y|) \quad (8)$$

$$\varepsilon = 3\sigma \sqrt{\frac{\ln D}{D}} \quad (9)$$

$$\hat{\sigma}^2 = \frac{D}{D-d} \cdot \frac{1}{D} \sum_{i=1}^D (y_i - \hat{y}_i)^2 \quad (10)$$

“Table 8”

4.4 Multiple Linear Regression (MLR)

The general form of a linear model is presented as Equation (11) where β is a matrix of all unknown correlation coefficients, X is a matrix of all inputs, and ε is an D -by-1 vector of random disturbances, where D is the number of observations. Using D training datasets, the working mechanism of MLR is to minimize the sum of the squared residuals by computing a closed-form expression for estimating the unknown parameter β .

$$Y = \beta X + \varepsilon \quad \text{where } \beta = [a \ b \ c \dots]^T \quad (11)$$

4.5 Response Surface Model (RSM)

The general form of a RSM can be also written as Equation (11). But, in addition to the linear terms, the matrix, X , must contain quadratic interaction terms and squared terms (or even higher order). Hence, the quadratic models of M1 and M2 are non-linear in nature.

4.6 Comparisons between ANN, GRNN, SVR, MLR and RSM

For benchmarking purposes, ANN was compared with GRNN, SVR, MLR, and RSM in terms of modeling accuracy. From Table 9, it can be seen that ANN, GRNN, and SVR can outperform MLR and RSM for M1. In terms of MAPE and SD, ANN may slightly perform better than SVR for M1. **It is also noted that GRNN performs slightly better than ANN in terms of MAPE, but with larger SD. It may imply that the performance of GRNN is less stable than that of ANN.** However, there is no single technique that can explicitly outperform other benchmarking methods for M2. Perhaps, M2 is relatively simple as RA is just a binary output while RT is a real number in M1. The results of two-sample t-test among ANN, SVR, MLR, and RSM are reported in Table 10. **For M1, the null hypothesis was accepted between ANN, GRNN, and SVR indicating that the performance of ANN was not significantly different from that of GRNN and SVR ($p < 0.05$). However, the null hypotheses were rejected between pairs of ANN-MLR, GRNN-MLR, and SVR-MLR, which means that the performance of ANN, GRNN and SVR was significantly different from that of MLR ($p < 0.01$). The same can be observed between pairs of ANN-RSM, GRNN-RSM, and SVR-RSM ($p < 0.01$).** Moreover, RSM was proved significantly better than MLR ($p < 0.01$). For M2, all null hypotheses were accepted indicating that the performances of all methods were statistically alike ($p < 0.05$). One possible explanation is that M2 is much simpler than M1, and the traditional regression methods would achieve comparable performances as the sophisticated methods (Wong and Chan, 2012), such as the ANN, GRNN, and SVR reported here.

“Tables 9-10”

4.7 Computation of Relative Influence of Variables

After ANN, GRNN, and SVR were deemed reliable, we used them to compute the relative influence (RI) of factors on RT (i.e. M1) and RA (i.e. M2), respectively. In this study, both change of mean square error (COM) and sensitivity analysis (SA) were applied to ANN, GRNN, and SVR for computing the RIs. The COM method is used to statistically rank the influence of the inputs (Sung, 1998). In brief, it measures the change in the prediction mean square error (MSE) of a regression model after an input is deleted. The prediction MSE with input n deleted is determined by Equation (12). The procedure is as follows: we re-train each regression model with $N-1$ inputs each time after the input n is deleted, and then compute the absolute difference between the MSE of the reduced model (MSE_n) relative to the MSE of the full model with all N inputs (MSE_{all}). We then rank the input whose deletion triggers the largest change in MSE as the most influential factor, since its exclusion from the full model causes the most deterioration in the prediction performance. The RI of factor n can then be quantified based on its proportion of the change induced relative to the total change in MSE induced by all factors, as defined by Equation (13). Another method, SA, analyzes the model responses by changing the inputs (Kewley et al., 2000). The procedure is as follows: we train each model by holding all inputs at their average values except input n , which varies through its entire range with L_n levels. If input n is important, it should produce a high variance (V_n) measured by Equation (14), where y_{n_i} is the model output when

the input n is held at its i -th level and other inputs are held at their average values, and \bar{y}_n is the average model output over all levels of input n . Thus, the RI of input n can be computed by Equation (15). To obtain a more robust estimation of the input influence, the RI of each factor is averaged over ten-fold training. It is noted that the number of levels of each input can be referred to Table 1 except RT. As RT is a continuous input, $L = 5$ is suggested (Cortez et al., 2006).

$$MSE_n = \frac{1}{D} \sum_{i=1}^D (A_i - E_i)^2 \quad (12)$$

$$RI_n = \frac{|MSE_n - MSE_{all}|}{\sum_{i=1}^N |MSE_i - MSE_{all}|} \quad (13)$$

$$V_n = \frac{1}{L_n - 1} \sum_{i=1}^{L_n} (y_{n_i} - \bar{y}_n)^2 \quad (14)$$

$$RI_n = \frac{V_n}{\sum_{i=1}^N V_i} \quad (15)$$

For benchmarking purposes, another method based on MLR is also used to compute the RIs. As recommended by Lindeman et al. (1980), this MLR-based method uses the squared semi-partial correlation of each predictor (input) as it is added to the model as the measure of importance (Johnson and LeBreton, 2011). In short, the RI of each input can be defined as the average contribution to R^2 across all possible orderings. This measure is meaningful as it considers the direct effect of an input, the total effect of all inputs, and the partial effect of all subsets of inputs (Johnson, 2000). Hence, totally five different ways of computing RIs were compared, i.e. ANN with COM method (ANN-COM), ANN with SA method (ANN-SA), **GRNN with COM method (GRNN-COM)**, **GRNN with SA method (GRNN-SA)**, SVR with COM method (SVR-COM), SVR with SA method (SVR-SA), and MLR-based method. Figures 5 and 6 show the RIs computed by different methods on M1 and M2 respectively, and the relevant result implications will be discussed in the following section. Since there is no well-developed method of computing the RIs based on RSM, it is excluded from the comparison. But, useful visualization of RSM in this study will be discussed in Section 5.3.

“Figures 5-6”

5. DISCUSSION

5.1 Response Time

Figure 5 shows that in predicting RT, the most influential factor was CD followed by RA. It can also be noted that all methods can produce similar RIs for M1, i.e. consensus can be reached. Hence, it is strongly believed that RT can be predicted by considering CD which contributes around 56%-71% (mean = 64%) of the overall influence. Since RA is also a measure of user performance, it cannot be freely adjusted or controlled. This is why RA is weighed only around 29%-44% (mean = 36%). Next, we need to determine how CD can be controlled to minimize RT via the design of experiment (DOE) approach. Figure 7 shows the main effect of CD and RA on RT. From the left part of Figure 7, it can be generally observed that the right-handed participants took shorter times (in ms) when reacting to an instruction using the horizontal control device (mean RT = 603 ms, SD = 225 ms, $n = 3474$) while they required longer reaction time using vertical (mean RT = 652 ms, SD = 276 ms, $n = 3502$) and rotary control devices (mean RT = 642 ms, SD = 199 ms, $n = 3510$). The right part of Figure 7 shows that if the right-handed participants reacted wrongly to the instruction, the reaction time was longer (mean RT = 661 ms, SD = 262 ms, $n = 1309$). On the other hand, the reaction time was shorter if the participants reacted correctly (mean RT = 628 ms, SD = 232 ms, $n = 9177$). This is true if the display-control mapping is incompatible, and users may need to exercise more mental effort, hence, leading to longer reaction times and higher probabilities of having the wrong reaction. However, RA is not a design variable in this study.

In summary, in order to minimize RT, a 2-way horizontal lever joystick must be used to execute the instructions which consist of L, R, U, D, C, and AC motions with displays oriented in the four cardinal directions.

“Figure 7”

5.2 Response Accuracy

Unlike M1, dissimilar RIs were computed by different methods on M2 as shown in Figure 6, i.e. consensus cannot be reached. Using ANN-COM, the importance of the inputs was ranked as $CD > RT > IN > DO$, which is the same as the ranking obtained by the method of Lindeman et al. (1980). Using ANN-SA, a slightly different ranking was obtained as $CD > IN > RT > DO$. **Using GRNN-COM, the ranking was the same as the one by ANN-COM. Similarly, using GRNN-SA, a different ranking was also generated as $CD, IN > DO > RT$.** Using the SVR-COM and SVR-SA methods, the rankings were quite different. Since MLR-based methods have been proved reliable and commonly adopted (Green et al., 1978; Johnson, 2000; Johnson, 1966), we believe that the RIs computed by ANN-COM and GRNN-COM must be more genuine than the two SVR-based methods. In this connection, for predicting RA, the most influential factor was CD followed by RT, IN, and DO. If CD, RT, and IN which together contribute 84%-89% of the overall influence on RA are considered only, DOE can also be applied. Figure 8 shows the main effect of CD and IN on RA. From the left part of Figure 8, it can be observed that 89.83% and 94.25% of the participants could react correctly to the instruction using horizontal ($n = 3474$) and rotary control devices ($n = 3510$) while only 78.47% of them could react correctly using the vertical control device ($n = 3502$). The association between RA and RT has been already discussed in Section 5.1 (see Figure 7). Similar to M1, RT is not a design variable in controlling RA, or vice versa. From the right part of Figure 8, 89.8%, 90.97%, and 90.3% of the participants could react correctly to L ($n = 1735$), R ($n = 1750$), and AC instructions ($n = 1732$) respectively. In contrast, only 84.49%, 83.38%, and 86.22% of the participants could react correctly to U ($n = 1760$), D ($n = 1745$), and C instructions ($n = 1764$) respectively. It is interesting to note that DO had little impact on RA. A possible explanation is as follows: given that the location of the control device was fixed, the right-handed participants would concentrate on how to react to an instruction using a particular type of control device. To observe an effect on the display oriented in a specific direction, participants only turned their head toward that display while keeping their hand on the control device. Hence, the association between the hand motion and the display was maintained regardless of the display direction. As changing the direction of display did not induce extra mental effort or unnecessary disruption to the participants, RA was found to be less sensitive to DO.

In summary, there are two suggestions in which RA can be maximized with four different display orientations: 1) a rotary or horizontal control device must be used; and 2) both U and D instructions must be avoided where the C instruction is a marginal case.

“Figure 8”

5.3 Visualization of RSM

RSM can also be used to visualize the impact of all inputs (CD, DO, and IN) on each of the two outputs (RT and RA), hence called response surface analysis (RSA). Figure 9 and Figure 10 reveal the impact of all inputs on RT and RA respectively. From Figure 9, it can be observed that RT is very sensitive to CD but it does not vary significantly with DO and IN. This finding is reinforced in Section 5.1 that CD is the most influential factor to RT. From Figure 10, it can be noted that RA varies significantly together with CD and IN. This observation is also reinforced in Section 5.2 reporting that CD and IN are more influential than IN in controlling RA. More importantly, the optimum response (output) can be addressed by searching for the best level of each input. In this study, we attempt to minimize RT and maximize RA. To do so, all the major recommendations can be visualized in Figure 9 and Figure 10.

“Figure 9-10”

6. CONCLUSIONS

A methodology of quantifying the causal connection between design variables and user performance was successfully applied. Using experimental data, the RI of each variable on each performance measure was

quantitatively computed. By employing correlation and partial correlation analyses, two relation graphs were constructed, one for each of the two key performance measures, RT and RA. Four regression methods were applied to quantify the connections in the relation graphs, and three of them were used to compute the RI of the design variables on RT and RA independently. **The comparison results suggest that ANN-based and GRNN-based methods are more reliable than SVR-based methods in computing the RI of design variables, although their modeling abilities are comparable (both are statistically better than MLR and RSM). Based on the RIs computed by ANN-COM and GRNN-COM, CD is a major design variable for controlling RT while CD and IN are two key variables for controlling RA.** As a result of our analysis, the best design option to optimize each of the measures is suggested. However, if the objective is to optimize both RT and RA simultaneously, the recommended settings for right-handed users using their right hand would be as follows: a 2-way horizontal control device should be used with only L, R, C, and AC instructions. In other words, if only one type of control device is allowed, a horizontal device must be the best option for users to execute 4 out of 6 distinct instructions. One may argue that a more versatile device should be used. But in fact, users may tend to make more mistakes (larger RT and lower RA) if they have more options given a number of distinct instructions and display directions. This is also one of our future research directions.

In general, the major contribution of this study is to suggest some useful quantitative methods in computing the relative influence among variables in the realm of display-control compatibility, such that system designers will be able to focus on some key variables rather than on all of them. **Also, the extraction of explanatory information from renowned black-box system such as the ANN and GRNN reported here is successfully demonstrated.** Generally speaking, our proposed method will be extremely helpful when there are too many design variables and even if some display-control relationships are highly implicit or cannot be defined quantitatively. However, there is no universal method that can absolutely identify the relative influence of inputs towards outputs in all cases (Sung, 1998; Wong and Chan, 2012). This suggests that our method must be tested on further problems, especially within the domain of ergonomics. In this regard, the authors plan to extend this research to different display-control relationships on user performance (including left-handed participants) with a variety of operating conditions, e.g. multiple controls and multiple displays. Other quantitative yet complex techniques will be examined and compared with ANN and SVR such as evidential reasoning. **Also, optimization algorithms will be incorporated into our methodology when determining the network parameters (P , σ , C , and ϵ) such that explanatory information can be extracted from a more reliable model.**

ACKNOWLEDGMENT

The work described in this paper was partially supported by a grant from City University of Hong Kong (SRG7004079).

REFERENCES

- Alenezi, A., Moses, S.A., and Trafalis, T.B. (2008). Real-time prediction of order flowtimes using support vector regression. *Computes & Operations Research*, 35, 3489-3503.
- Bi, L., Tsimhoni, O., and Liu, Y. (2011). Using the support vector regression approach to model human performance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 41, 410-417.
- Bishop, C.M. (1995). *Neural networks for pattern recognition*. Clarendon Press, Oxford.
- Brebner, J. and Sandow, B. (1976). The effect of scale side on population stereotype. *Ergonomics*, 19, 571-580.
- Burgess-Limerick, R., Krupenia, V., Wallis, G., Pratim-Bannerjee, A. and Steiner, L. (2010). Directional control-response relationships for mining equipment. *Ergonomics*, 53, 748-757.
- Burgess-Limerick, R., Zupanc, C.M. and Wallis, G. (2012). Directional control-response compatibility of joystick steered shuttle cars. *Ergonomics* (in press).
- Cao, Q., Leggio, K.B., Schniederjans, M.J. (2005). A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market. *Computers and Operations Research*, 32, 2499-2512.
- Chan, A.H.S. and Chan, W.H. (2006). Movement compatibility for circular display and rotary controls positioned at peculiar positions. *International Journal of Industrial Ergonomics*, 36, 737-745.
- Chan, W.H. and Chan, A.H.S. (2003). Movement compatibility for rotary control and circular display-computer simulated test and real hardware test. *Applied Ergonomics*, 34, 61-67.
- Chan, W.H. and Chan, A.H.S. (2007). Strength and reversibility of movement stereotypes for level control and circular display. *International Journal of Industrial Ergonomics*, 37, 233-244.

- Chan, W.H. and Chan, A.H.S. (2008b). Strength and reversibility of stereotypes for rotary control with linear scale. *Perceptual and Motor Skills*, 106, 341-353.
- Chan, W.W.H. and Chan, A.H.S. (2008a). Movement compatibility for two-dimensional level control and digital counter. *IEEE Transactions on Systems, Man, and Cybernetics*, 38, 528-533.
- Chen, K.Y. (2007). Forecasting systems reliability based on support vector regression with genetic algorithms. *Reliability Engineering and System Safety*, 92, 423-432.
- Chen, K.Y. and Wang, C.H. (2007). Support vector regression with genetic algorithms in forecasting tourism demand. *Tourism Management*, 28, 215-226.
- Cherkassky, V. and Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, 17, 113-126.
- Cherkassky, V., Shao, X., Mulier, F., and Vapnik, V. (1999). Model complexity control for regression using VC generalization bounds. *IEEE Transactions on Neural Networks*, 10, 1075-1089.
- Chua, R., Weeks, D.J., Ricker, K.L., Poon, P. (2001). Influence of operator orientation on relative organizational mapping and spatial compatibility. *Ergonomics*, 44, 751-765.
- Coren, S. (1993). Measurement of handedness via self-report: The relationship between brief and extended inventories. *Perceptual and Motor Skills*, 76, 1035-1042.
- Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47, 547-533.
- Cortez, P., Portelinha, M., Rodrigues, S., Cadavez, V., and Teixeira, A. (2006). Lamb meat quality assessment by support vector machines. *Neural Processing Letters*, 24, 41-51.
- Courtney, A.J. (1992). Display-control stereotypes for multicultural user systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 22, 681-687.
- Courtney, A.J. (1994a). Hong Kong Chinese direction-of-motion stereotypes. *Ergonomics*, 37, 417-426.
- Courtney, A.J. (1994b). The effect of scale-side, indicator type, and control plane on direction-of-turn stereotypes for Hong Kong Chinese subjects. *Ergonomics*, 37, 865-877.
- Deng, W.J. and Pei, W. (2009). Fuzzy neural based importance-performance analysis for determining critical service attributes. *Expert Systems with Applications*, 36, 3774-3784.
- Deng, W.J., Chen, W.C., and Pei, W. (2008). Back-propagation neural network based importance-performance analysis for determining critical service attributes. *Expert Systems with Applications*, 34, 1115-1125.
- Enke, D. and Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with Applications*, 29, 927-940.
- Garson, G.D. (1991). Interpreting neural-network connection weights. *AI expert*, 6, 47-51.
- Green, P.E., Carroll, J.D., and DeSarbo, W.S. (1978). A new measure of predictor available importance in multiple regression. *Journal of Marketing Research*, 15, 356-360.
- Haidar, A.M.A., Mustafa, M.W., Ibrahim, F.A.F., and Ahmed, I.A. (2011). Transient stability evaluation of electrical power system using generalized regression neural networks. *Applied Soft Computing*, 11, 3558-3570.
- Hoffman, E.R. (1997). Strength of component principles determining direction of turn stereotypes-linear displays with rotary controls. *Ergonomics*, 40, 199-222.
- Howes, P. and Crook, N. (1999). Using input parameter influences to support the decisions of feedforward neural networks. *Neurocomputing*, 24, 191-206.
- Hu, M. Y. and Tsoukalas, C. (2003). Explaining consumer choice through neural networks: the stacked generalization approach. *European Journal of Operational Research*, 146, 650-660.
- Huang, Z., Chen, H., Hsu, C-J., Chen, H-W., and Wu, S. (2004). Credit rating analysis with support vector machine and neural networks: a market comparative study. *Decision Support Systems*, 37, 543-558.
- International Organization for Standardization, 1998. ISO 9241-5:1998. Ergonomics requirements for office work with visual display terminals (VDTs) – Part 5: Workstation layout and postural requirements.
- Johnson, J.W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, 35, 1-19.
- Johnson, J.W. and LeBreton, J.M. (2011). History and use of relative importance indices in organizational research. *Organizational Research Methods*, 7, 238-257.
- Johnson, R.M. (1966). The minimal transformation to orthonormality. *Psychometrika*, 31, 61-66.
- Kang, H.G. and Seong, P. H. (2001). Information theoretic approach to man-machine interface complexity evaluation. *IEEE Transactions on Systems, Man, and Cybernetics*, 31, 163-171.
- Kawato, M., Isobe, M., Uno, Y., and Suzuki, R. (1990). Trajectory formation of arm movements by cascade neural network model based on minimum torque-change criterion. *Biological Cybernetics*, 62, 275-288.
- Kewley, R.H., Embrechts, M.J., and Breneman, C. (2000). Data strip mining for the virtual design of pharmaceuticals with neural networks. *IEEE Transactions on Neural Networks*, 11, 668-679.

- Kim, B., Kwon, S., and Kim, D. H. (2010). Optimization of optical lens-controlled scanning electron microscopic resolution using generalized regression neural network and genetic algorithm. *Expert Systems with Application*, 37, 182-186.
- Kim, H.J. and Shin, K.S. (2007). A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets. *Applied soft computing*, 7, 569-576.
- Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., and Aulagnier, S. (1996). Application of neural networks to modeling nonlinear relationships in ecology. *Ecological Modelling*, 90, 39-52.
- Lindeman, R.H., Merenda, P.F., and Gold, R.Z. (1980). *Introduction to bivariate and multivariate analysis*. Glenview, IL: Scott, Foresman and Company.
- Lu, C.J., Lee, T.S., Chiu, C.C. (2009). Financial time series forecasting using independent component analysis and support vector regression, *Decision Support Systems*, 47, 115-125.
- Makiguchi, M., Tokunaga, H. and Kanamori, H. (2003). A human factors study of switches installed on automotive steering wheel. *JSAE Review*, 24, 341-346.
- McCarley, J.S. and Wickens, C.D. (2005). Human factors implications of UAVs in the national airspace. Technical Report AHFD-05-05/FAA-05-01. Urbana, IL: Aviation Human Factors Division, Institute of Aviation, University of Illinois at Urbana-Champaign.
- Olden, J. D. and Jackson, D. A. (2002). Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, 154, 135-150.
- Olden, J. D., Joy, M. K., and Death, R. G. (2004). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*, 178, 389-397.
- Rigotti, C., Cerveri, P., Andreoni, G., Pedotti, A., and Ferrigno, G. (2001). Modeling and driving a reduced human mannequin through motion captured data: a neural network approach. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 31, 187-193.
- Sen, R.N. and Das, S. (2000). An ergonomics study on compatibility of controls overhead cranes in a heavy engineering factory in West Bengal. *Applied Ergonomics*, 31, 179-184.
- Smola, A.J. and Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14, 199-222.
- Specht, D. (1991). A general regression neural network. *IEEE Transactions on Neural Networks*, 2, 568-576.
- Sung, A.H. (1998). Ranking importance of input parameters of neural networks. *Expert Systems with Applications*, 15, 405-411.
- Tsaur, S.H., Chiu, Y.C., and Huang, C.H. (2002). Determinants of guest loyalty to international tourist hotels – a neural network approach. *Tourism Management*, 23, 397-405.
- Uno, Y., Kawato, M., and Suzuki, R. (1989). Formation and control of optimal trajectory in human arm movement. *Biological Cybernetics*, 61, 89-101.
- Vapnik, V.N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10, 988-999.
- Vapnik, V.N. (2000). *The nature of statistical learning theory*. New York: Springer.
- Wang, Y.-M. and Elhag, T.M.S. (2007). A comparison of neural network, evidential reasoning and multiple regression analysis in modeling bridge risks. *Expert Systems with Application*, 32, 336-348.
- White, H. (1990). Connectionist nonparametric regression: multilayer feedforward networks can learn arbitrary mappings. *Neural Networks*, 3, 535-549.
- Wong, T.C. and Chan, Alan H.S. (2012). A study of the impact of different direction-of-motion stereotypes on response time and response accuracy using neural network. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 42, 1077-1087.
- Wong, T.C., Law, Kris M.Y., Yau, Hon K., and Ngan, S.C. (2011). Analyzing supply chain operation models with the PC-algorithm and the neural network. *Expert Systems with Applications*, 38, 7526-7534.
- Worringham, C.J. and Beringer, D.B. (1998). Directional stimulus-response compatibility: a test of three alternative principles. *Ergonomics*, 41, 864-880.
- Yoon, Y., Guimaraes, T., and Swales, G. (1994). Integrating artificial neural networks with rule-based expert systems. *Decision Support Systems*, 11, 497-507.
- Yu, L., Wang, S., and Lai, K.K. (2008). Credit risk assessment with a multistage neural network ensemble learning approach. *Expert Systems with Applications*, 34, 1434-1444.
- Zhu, G., Huang, Q., Xu, C., Xing, L., Gao, W., and Yao, H. (2007). Human behavior analysis for highlight ranking in broadcast racket sports video. *IEEE Transactions on Multimedia*, 9, 1167-1182.
- Zupanc, C.M., Burgess-Limerick, R.J., and Wallis, G. (2007). Performance consequences of alternating directional control-response compatibility: Evidence from a coal mine shuttle car simulator. *Human Factors*, 49, 629-636.

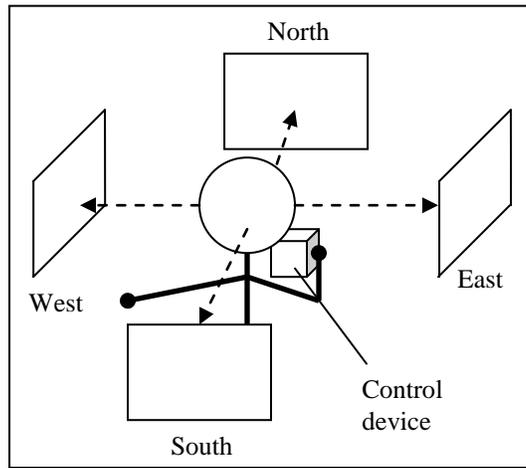


Figure 1: Experimental configuration

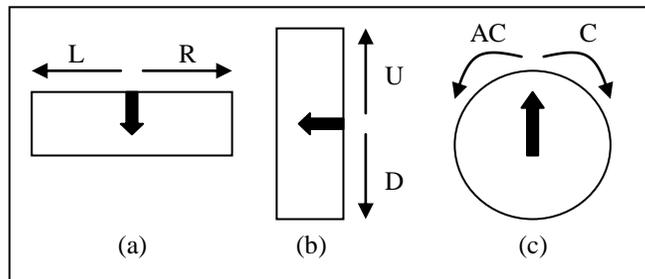


Figure 2: Display scales

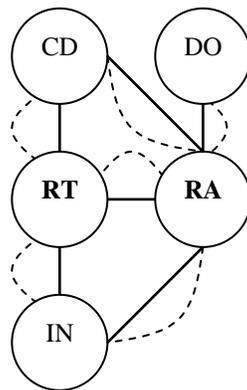


Figure 3: Initial relation graph

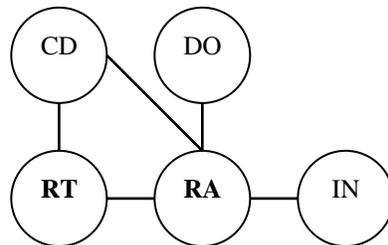


Figure 4: Revised relation graph

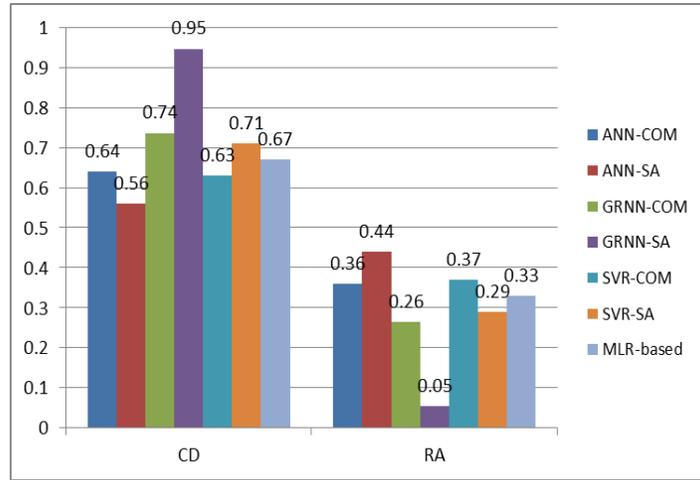


Figure 5: RIs computed by different methods on M1

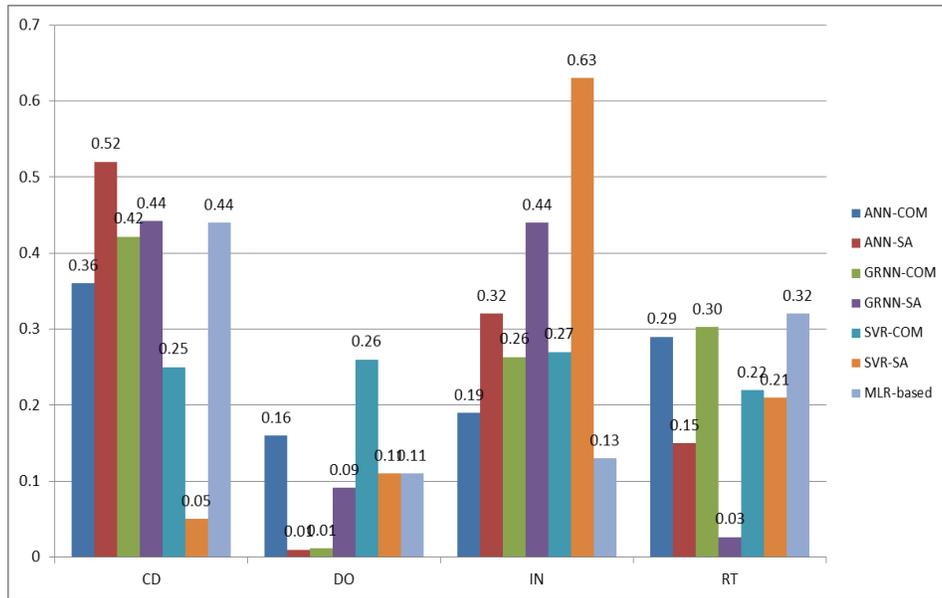


Figure 6: RIs computed by different methods on M2

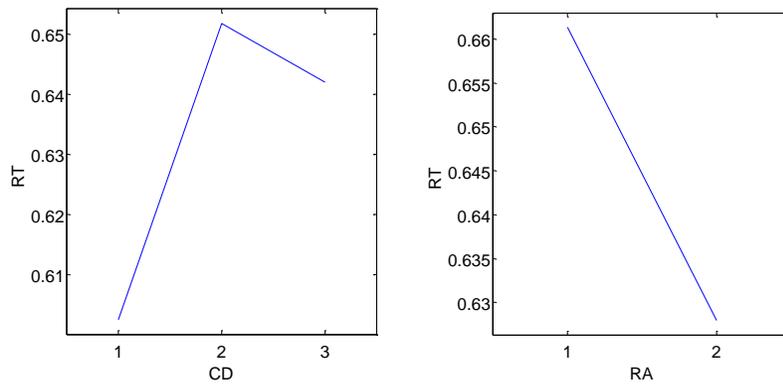


Figure 7: Main effect of CD and RA on RT in seconds

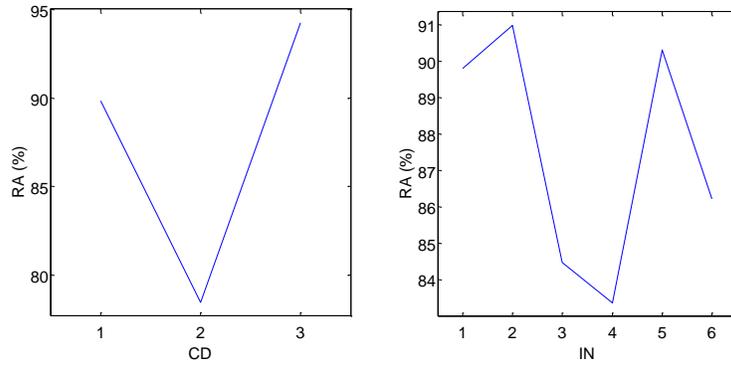


Figure 8: Main effect of CD and IN on RA in %

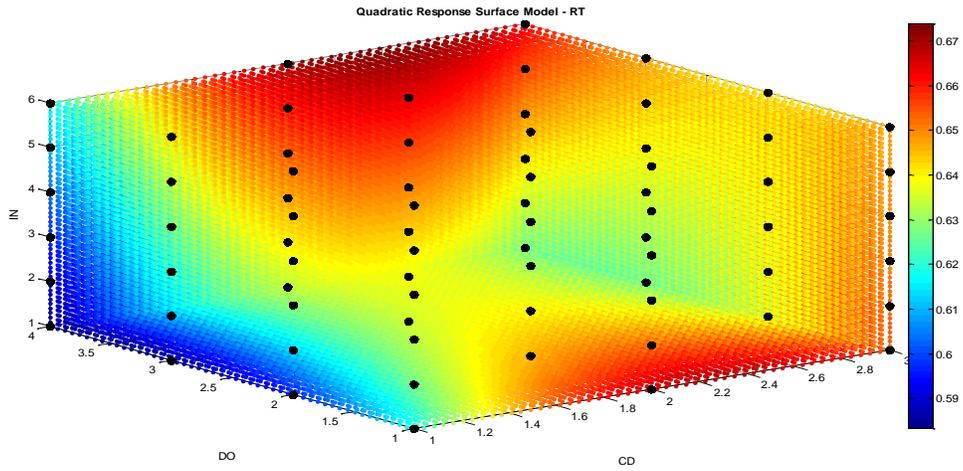


Figure 9: Impact of all inputs on RT

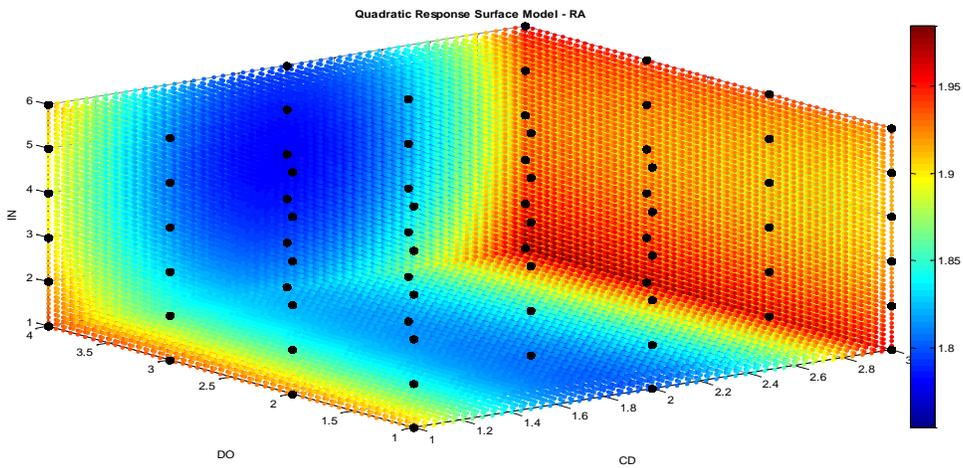


Figure 10: Impact of all inputs on RA

Table 1: Levels of controllable inputs and measurable outputs

Variables	Factors	Levels
Inputs	CD	1 – Horizontal 2 – Vertical 3 – Rotary
	DO	1 – North (N) 2 – East (E) 3 – South (S) 4 – West (W)
	IN	1 – Left (L) 2 – Right (R) 3 – Up (U) 4 – Down (D) 5 – Anti-clockwise (AC) 6 – Clockwise (C)
Outputs	RP	1 – Left (L) 2 – Right (R) 3 – Up (U) 4 – Down (D) 5 – Anti-clockwise (AC) 6 – Clockwise (C)
	RT	Not applicable
	RA	1 – False 2 – True

Table 2: Stereotypes of horizontal control

DO	IN	1-L		2-R		3-U		4-D		5-AC		6-C	
	RP	1-L	2-R	1-L	2-R	1-L	2-R	1-L	2-R	1-L	2-R	1-L	2-R
North		100	0	0	100	32.88	67.12	75.51	24.49	91.72	8.28	0	100
East		100	0	0	100	17.24	82.76	68.57	31.43	100	0	0	100
South		97.24	2.76	1.35	98.65	34.72	65.28	69.66	30.34	97.10	2.90	0	100
West		100	0	0	100	29.73	70.27	72.97	27.03	100	0	0	100

Table 3: Stereotypes of vertical control

DO	IN	1-L		2-R		3-U		4-D		5-AC		6-C	
	RP	3-U	4-D	3-U	4-D	3-U	4-D	3-U	4-D	3-U	4-D	3-U	4-D
North		35.14	64.86	65.52	34.48	97.30	2.70	5.52	94.48	27.70	72.30	54.05	45.95
East		33.33	66.67	66.67	33.33	97.30	2.70	3.55	96.45	24.32	75.68	53.38	46.62
South		20.27	79.73	80.69	19.31	94.59	5.41	5.52	94.48	21.62	78.38	64.79	35.21
West		28.77	71.23	80.41	19.59	97.30	2.70	0.69	99.31	23.97	76.03	61.11	38.89

Table 4: Stereotypes of rotary control

DO	IN	1-L		2-R		3-U		4-D		5-AC		6-C	
	RP	5-AC	6-C	5-AC	6-C	5-AC	6-C	5-AC	6-C	5-AC	6-C	5-AC	6-C
North		100	0	0	100	10.88	89.12	83.78	16.22	95.89	4.11	0	100
East		100	0	0	100	19.18	80.82	82.31	17.69	100	0	0	100
South		97.30	2.70	0	100	13.70	86.30	78.38	21.62	98.61	1.39	0	100
West		100	0	0	100	15.07	84.93	84.93	15.07	100	0	0	100

Table 5: Correlation analysis

	Controllable Inputs			Major Outputs	
	CD	DO	IN	RT	RA
CD	-	-0.0015	0.0012	0.0679**	0.0550**
DO		-	-0.0044	-0.0155	0.0289*
IN			-	0.0198#	-0.0310*
RT				-	-0.0466**

Note: ** p=0; * p<0.01; # p<0.05

Table 6: Average cross-validation results using ANN with varying P

Model	P	3	4	5	6	7	8	9	10
M1	MAPE	9.88	9.86	9.85	9.78	9.85	9.91	9.82	9.86
	SD	0.36	0.39	0.36	0.34	0.40	0.34	0.43	0.36
	MAX	10.57	10.57	10.55	10.42	10.51	10.60	10.48	10.52
	MIN	9.38	9.34	9.35	9.39	9.41	9.42	9.26	9.21
M2	PC	87.3	87.52	87.51	87.51	87.51	87.42	87.42	87.42

Table 7: Average cross-validation results using GRNN with varying σ

Model	σ	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
M1	MAPE	9.78	9.76	9.76	9.72	9.72	9.70	9.69	9.74	9.73	9.76	9.76
	SD	0.44	0.49	0.49	0.50	0.52	0.50	0.52	0.52	0.51	0.54	0.55
	MAX	10.31	10.40	10.47	10.38	10.50	10.39	10.40	10.38	10.38	10.37	10.37
	MIN	9.04	8.86	8.90	8.81	8.80	8.81	8.78	8.79	8.78	8.79	8.74
M2	PC	87.51	87.51	87.51	87.51	87.51	87.51	87.51	87.51	87.51	87.51	87.51

Table 8: Average cross-validation results using SVR with varying λ

Model	λ	0.1	0.5	2^{-7}	$1/2\sigma^2$
M1	MAPE	9.80	9.84	9.84	9.84
	SD	0.56	0.57	0.57	0.58
	MAX	10.44	10.47	10.47	10.46
	MIN	8.66	8.63	8.62	8.62
M2	PC	87.51	87.51	87.51	87.51

Table 9: Comparison results between ANN, GRNN, SVR, MLR, and RSM on M1 and M2

Model	Performance	ANN	GRNN	SVR	MLR	RSM
M1	MAPE (%)	9.78	9.69	9.80	11.89	10.58
	SD	0.34	0.52	0.56	0.57	0.38
M2	PC (%)	87.52	87.51	87.51	87.51	87.51

Table 10: Two-sample T-test between ANN, GRNN, SVR, MLR, and RSM on M1 and M2

Model	Methods	ANN	GRNN	SVR	MLR	RSM
M1	ANN	-	Accept	Accept	Reject	Reject
	GRNN		-	Accept	Reject	Reject
	SVR			-	Reject	Reject
	MLR				-	Reject

	RSM					-
M2	ANN	-	Accept	Accept	Accept	Accept
	GRNN		-	Accept	Accept	Accept
	SVR			-	Accept	Accept
	MLR				-	Accept
	RSM					-