Extracting the contribution of independent variables in neural network models: A new approach to handle instability

By: Juan de Oña & Concepción Garrido

Extracting the contribution of independent variables in neural network models: a new approach to handle instability

Juan de Oña[a] and Concepción Garrido

TRYSE Research Group. Department of Civil Engineering, University of Granada, ETSI Caminos, Canales y Puertos, c/ Severo Ochoa, s/n, 18071 Granada (Spain),

[a] Corresponding author. Phone: +34 958 24 99 79, Fax: +34 958 24 61 38, jdona@ugr.es

ABSTRACT

One of the main limitations of Artificial Neural Networks (ANN) is their high inability to know in an explicit way the relations established between explanatory variables (input) and dependent variables (output). This is a major reason why they are usually called "black boxes". In the last few years several methods have been proposed to assess the relative importance of each explanatory variable. Nevertheless, it has not been possible to reach a consensus on which is the best-performing method. This is largely due to the different relative importance obtained for each variable depending on the method used. This importance also varies with the designed network architecture and/or with the initial random weights used to train the ANN. This paper proposes a procedure that seeks to minimize these problems and provides consistency in the results obtained from different methods. Essentially, the idea is to work with a set of neural networks instead of a single one. The proposed procedure is validated using a database collected from a customer satisfaction survey which was conducted on the public transport system of Granada (Spain) in 2007. The results show that, when each method is applied independently, the variable's importance rankings are similar and, in addition, coincide with the hierarchy established by researchers who have applied other techniques.

Keywords: instability; neural networks; black box; variables contribution´s methods; importance ranking

1. INTRODUCTION

ANN are information processing systems based on the biological behavior of the human brain and used in a growing number of multiple research fields. The strength of ANN compared to other techniques is their high capacity for classification, prediction and failure tolerance (Martín del Brío and Sanz, 2006).

ANN are not based on a predefined equation or formula, but on their capacity to capture the information inherent to the data submitted during the training process. They create an architecture whose parameters are able to provide correct answers when some new cases are presented. These parameters are the key to their knowledge (Palmer and Montaño, 2002a). This singular way of learning allows them to capture highly non-linear (Watts and Worner, 2008) and complex (Mohammadipour and Alavi, 2009) relations, but prevents an explicit explanation of how explanatory variables (input) and dependent variables (output) are related. However, it is possible to achieve this goal with classical statistical techniques (Azadeh et al., 2011). Therefore ANN are included in the group of data mining techniques called "black boxes" (Cortez and Embrechts, 2013), because for a given phenomenon (output), it is very difficult to know the relative importance of each variable (input).

Considering this problem, several methods have been proposed to determine the contribution of each independent variable in ANN models; many methods from the family of sensitivity analyses (SA), which basically changes input values and checks what happens in the output, and others specific NN methods have been developed by the researchers. SA methods perform a pure black-box approach over a data-driven model, which can be NN or other method, such as SVM. In contrast, specific NN methods can only be applied to one typology of NN (the multilayer perceptron), so that they are not universal input relevance methods. Sung (1998) applied three methods (sensitivity analysis (Zurada et al., 1994; Engelbrecht et al., 1995), fuzzy curves (Lin and Cunningham, 1995) and change of Mean Square Error (MSE) (He et al., 1997)) to a database related to petroleum engineering and compared the results obtained. It was concluded that the fuzzy curves method performs better than the other two. Olden and Jackson (2002) described the neural interpretation diagram (Özesmi and Özesmi, 1999), Garson´s algorithm (Garson, 1991; Goh, 1995) and sensitivity analysis (Lek et al, 1995; 1996a, 1996b) methods, and proposed a new one called randomization approach. After applying these methods to a database related to ecology, it was observed that the results were different depending on the method used. Palmer and Montaño (2002b) highlighted the problems of the methods studied so far, both those based on the weights of connections and those based on sensitivity analysis. They argued that several previous studies had demonstrated that the former group of methods is not effective (Garson, 1991; Rzempoluk, 1998; Hunter et al., 2000), and that the latter presents some problems depending on the qualitative or quantitative nature of the variables. Therefore, they suggested a new approach called numeric sensitivity analysis (NSA) (Palmer and Montaño, 2003), which determines the relation between each input and output variable through the slope, and without taking into account the qualitative or quantitative nature of the variables. Gevrey et al. (2003) analyzed and compared seven methods: partial derivatives (Dimopoulos et al., 1995), Garson (Garson, 1991; Goh, 1995), perturb (Yao et al., 1998; Scardy and Harding, 1999), profile (Lek et al. 1995; 1996a; 1996b), and stepwise with some of its variants (Sung, 1998). They used an empirical ecological database and concluded that the partial derivatives method was the best, while the classical stepwise performed the worst. Olden et al. (2004) continued the work of Gevrey et al. (2003) and applied the same seven methods plus a new approach called connection weights (Olden and Jackson, 2002) to a simulated database. The connection weights method provided the best results.

Gevrey et al. (2006) introduced a new variation of the partial derivatives method (PaD), called PaD2, with the aim of analyzing the joint contribution of every possible pairwise combination of variables. They argued that in nature variables normally interact with each other, so when a variable is modified the remaining variables also change. Kemp et al. (2007) implemented the Holdback Input Randomization (HIPR) method, based on the random alteration of NN input parameters. They applied this method to ecological complex systems and the results obtained were as good as those extracted from the connection weights method. Yeh and Cheng (2010) provided a new point of view to the contribution of variables in NN by introducing a method that considers not only linear effects (first order derivatives) between the studied variables but also curvature effects (second order derivatives). Cortez and Embrechts (2013) introduced three new sensitivity analysis (SA) methods: data-based SA (DSA), Monte-Carlo SA (MSA) and cluster-based SA (CSA), and they compared them with two other existing methods: one-dimensional SA (1D-SA) (Kewley et al., 2000) and global SA (GSA) (Cortez and Embrechts, 2011). In addition, they developed some new approaches to determine the relative importance of variables and also pairs of input variables with sensitivity analysis methods.

Commonly, the first step is to select an optimal NN architecture, and the second step consists in applying several variable contribution methods. But it is very difficult to choose the optimal NN model due to various factors such as random initialization of connection weights, multiple possible network architectures and learning algorithms that converge towards local minimums in a complex error surface. Cao and Qiao (2008) underlined this issue and suggested that sensitivity analysis methods should be applied not to a single NN, but to a set of good-performing NN. The authors proposed a new application called neural network committee (NNC). Paliwal and Kumar (2011) also referred to the high variability of weights before starting the training and proposed an approach named interquartile range, in which the network is trained a number of times, and the first and third quartiles of the weight distribution are used to determine the relative importance of each variable.

However, despite the variety of methods studied, there is no general consensus on which model is the best for determining the contribution of variables. When applying several methods to optimal or sub-optimal NN architecture, the importance ranking of variables differs from method to method, indicating their inherent instability.

This paper intends to handle the instability problems derived from these methods, and suggests a new systematic application of the existing methods in order to obtain similar results of the importance ranking of variables, independently of the method applied.

The database used in this paper is based on a customer satisfaction survey developed by the Transport Consortium of the Granada Metropolitan Area (Spain) in 2007. De Oña et al. (2012; 2013) have analyzed this database with other methodological approaches (e.g., decision trees and structural equation approaches) in order to identify the most important variables that contribute to the perception of service quality in a public transportation service.

The paper is structured in five sections. Section 2 describes artificial neural networks, the methods used to determine the contribution of variables, the methodology followed and the database used in this study. Sections 3 and 4 continue with the results and discussion. The paper concludes with a summary and directions for future research.

2. NEURAL NETWORKS AND METHODS

2.1. Neural Networks (NNs)

The multilayer perceptron (MLP) is a widely used NN typology, introduced by Werbos (1974) and further developed and popularized by Rumelhart and McClelland (1986). The multilayer feed-forward NN has been used in approximately 70% of all ANN studies (Gedeon et al., 1995). They are so successful because several research groups (Funahashi 1989; Hornik et al., 1989) have mathematically demonstrated that a MLP neural network with a single hidden layer is a universal function approximator.

A gradient-descent supervised learning algorithm with a learning rate of 0.1 and a momentum of 0.9 were used to train the NN. This algorithm trains the NN by iteratively updating the synaptic weight values until the error function reaches a local minimum. The learning rate and momentum values help to accelerate the convergence (Rumelhart et al., 1986; Hagan et al., 1996). The weights were initialized before each training with small random values and a number of 20,000 epochs was considered.

The database was randomly divided into training, validation and test sets, in a 70:15:15 ratio.

A wide range of NNs was trained, all of them characterized by a three-layer architecture: an input layer with I neurons (one neuron per input variable), a hidden layer with H neurons ($H \in [1, N]$) and an output layer with J neurons (one neuron per output variable). The neurons were activated using logarithmic sigmoidal transfer functions in all layers.

2.2. Methods for determining the contribution of variables

The methods selected in this study to determine the relative importance of variables in a NN model have been proposed and applied by numerous authors in several research fields (Olden and Jackson, 2002; Gevrey et al., 2003): perturb, profile, connection weights and partial derivatives. The first two (perturb and profile) are pure SA methods, while the other two (connection weights and partial derivatives) are not.

2.2.1. Perturb method

This method is based on the principle of disturbing or introducing noise to one of the inputs while the remaining variables keep their original values. Afterwards, the mean square error (MSE) between the outputs obtained before and after the perturbation are compared (Yao et al., 1998; Scardy and Harding, 1999).

A noise δ was progressively applied to each variable in five steps: 10%, 20%, 30%, 40% and 50% of its original value. Thus, the variable x_i changes its values to $x_i = x_i + \delta$ due to the perturbation.

2.2.2. Profile method

This method analyses the evolution of each input along a scale or range of values, while the remaining variables keep their values fixed (Lek et al., 1995; 1996a; 1996b).

Each predictor variable x_i takes 11 different values resulting from the division of the range, between its minimum and maximum value, into 10 equal intervals. Furthermore, all variables except one are initially fixed at their minimum value, and then successively at their first quartile, median, third quartile and maximum value. Thus, 5 values of the response variable are obtained for each of the 11 values adopted by x_i, and the median of those 5 values is calculated. Finally, a curve with the profile of variation is obtained for every variable.

2.2.3. Connection weights method

This method determines the relative importance of the predictor variables of the model as a function of the NN synaptic weights, according to the mathematical expression (Olden and Jackson, 2002):

$$R_{ij} = \sum_{H=1}^{k} W_{ik} \cdot W_{kj} \qquad (1)$$

where R_{ij} is the relative importance of the variable x_i with respect to the output neuron j, H is the number of neurons in the hidden layer, W_{ik} is the synaptic connection weight between the input neuron i and the hidden neuron k, and W_{kj} is the synaptic weight between the hidden neuron k and the output neuron j.

2.2.4. Partial derivatives method

This method analyses the first order effects of the predictor variables of the model with respect to the output variable, using all available training data (Dimopoulos et al., 1995).

The output provided by a neuron of the hidden layer of a MLP neural network with sigmoidal activation functions is given by the following equations:

$$\mathrm{h}_k = \frac{1}{(1+\mathrm{e}^{-\mathrm{net}_k})} \tag{2}$$

$$\mathrm{net}_k = \sum\nolimits_{\mathrm{i}} \mathrm{W}_{ik} \cdot x_{\mathrm{i}} - \theta_k \tag{3}$$

where h_k is the output of the neuron k of the hidden layer, x_i is the value of the predictor variable of the considered input layer, W_{ik} is the connection weight between the predictor variable x_i and the neuron k of the hidden layer, and θ_k is the bias of the neuron k of the hidden layer.

The output of a neuron of the output layer is given by the following expressions:

$$\mathrm{y}_j = \frac{1}{(1+\mathrm{e}^{-\mathrm{net}_j})} \tag{4}$$

$$\mathrm{net}_j = \sum\nolimits_{\mathrm{j}} \mathrm{W}_{kj} \cdot \mathrm{h}_k - \theta_j \tag{5}$$

where W_{kj} is the connection weight between the neuron k of the hidden layer and the neuron j of the output layer, θ_j is the bias of the output neuron j and y_j is the output of the neuron j of the output layer.

The expression that relates the variation of the output values y_j with respect to the variation of the predictor variable x_i is obtained through application of the chain rule:

$$\frac{\mathrm{dy}_j}{\mathrm{dx}_i} = \sum\nolimits_{\mathrm{k}} \frac{\mathrm{dy}_j}{\mathrm{dnet}_j} \cdot \frac{\mathrm{dnet}_j}{\mathrm{dh}_k} \cdot \frac{\mathrm{dh}_k}{\mathrm{dnet}_k} \cdot \frac{\mathrm{dnet}_k}{\mathrm{dx}_i} = \sum\nolimits_{\mathrm{k}} \mathrm{f'}_j \cdot \mathrm{W}_{kj} \cdot \mathrm{f'}_k \cdot \mathrm{W}_{ik} \tag{6}$$

$$\mathrm{f'}_j = \mathrm{y}_j \cdot (1 - \mathrm{y}_j) \tag{7}$$

$$\mathrm{f'}_k = \mathrm{h}_k \cdot (1 - \mathrm{h}_k) \tag{8}$$

The sensitivity value of every variable x_i is given by the expression:

$$\mathrm{L}_i = \frac{\sum_{\mathrm{P}} \frac{\mathrm{dy}_j}{\mathrm{dx}_i}}{\mathrm{P}} \tag{9}$$

where P is the total number of training examples.

2.3. Methodology

The methodology proposed in this paper focuses on working with sets of NNs instead of a single NN. Every set is composed of a series of NNs with the same architecture, which are trained using an identical learning algorithm, activation functions, momentum value and learning ratio. NNs of the same set only differ in the initial random weight values considered in each training process. Once the NNs have been trained, the above four methods are applied to every one of them, and therefore a ranking of relative importance is obtained for each NN and for each method. Due to the instability of the results when a single method is used, this methodology proposes an approach based on calculating the ranking of relative importance for each method as a function of the average importance values obtained from every NN in the set.

MATLAB software was used to develop the NNs (Beale et al., 2007).

The sequence of steps followed to develop this methodology was:

Step 1. Train every one of the H NN architectures, with H Є [1, N] neurons in the hidden layer, M number of times, and using different random initial weights for each training. Thus, NxM trained ANNs are obtained..

Step 2. Determine the performance or capacity of generalization of the NxM trained NNs through the E error metric.

Step 3. Calculate the mean E and its standard deviation values for each one of the NN architectures, and select the NN architecture that reaches the global minimum mean E value.

Step 4. Apply the methods of variable contribution to every one of the M NNs of the selected architecture.

Step 5. For every method applied, calculate the average value of the M NNs for every one of I predictor variables.

Step 6. Determine the ranking of importance of I variables considered in the study based on the values obtained in Step 5 for every applied method.

Step 7. Compare the results obtained for each method.

2.4. Data

The database used in this study was obtained through a customer service quality survey performed on public bus users by the Granada Area Transport Consortium in 2007. This Consortium was created to coordinate and organize the transit bus service of the Metropolitan Area of Granada (Spain).

858 surveys were conducted at the bus stops of different lines, with the aim of measuring the user satisfaction level regarding the service quality provided through 12 variables. The answers to these variables were scored from 0 to 10, as shown in Table 1.

(Table 1 here)

In this case of study, the unitary value of the score obtained for each variable has been considered. That is, a range of values in the interval [0,1] has been used as input values for every variable, instead of using the original interval [0,10]. This translation allows to adapt them for subsequent treatment in the NN (Masters, 1993; Martín del Brío and Sanz, 2006), since the limits of the value range of every variable directly coincide with the upper and lower limits of the sigmoidal activation functions used in NN models.

3. RESULTS

In this study, it has been considered that N=30 and M=50, which came to a total of 1,500 networks, and the E error has been determined through the MAPE value, calculated according to the expression (Delen et al., 2006):

$$MAPE = \frac{1}{T} \cdot \sum_{i=1}^{T} abs(\frac{Actual\ value\ i - Set\ point\ value\ i}{Set\ point\ value\ i}) \tag{10}$$

where T is the total number of considered cases in the test stage.

Figure 1 shows the mean and the standard deviation values of MAPE obtained for every NN architecture. The MLP neural network with 6 neurons in the hidden layer reaches the minimum mean MAPE value among all trained networks. This network architecture, along with its 50 MLP trained networks, is selected to analyse the outcomes of the four methods considered in this paper.

(Figure 1 here)

Figures 2 to 5 show the relative importance of the 12 independent variables of the 50 networks used, obtained after applying the four methods. A wide variability of the importance values has been observed; hence every variable ranking varies greatly, depending not only on the method applied, but also on the different random initial values used for a same method. Some other researchers (Zhou et al., 2002; Cao and Qiao, 2008) have already set out this problem and suggested the possibility of working with NN sets to control the instability problems derived from sensitivity analysis. This supports the idea of working not with a single NN but with a set of them, as we are proposing in this paper.

The profile and perturb methods are able to enclose the relative importance of certain variables in a narrower range, as in the case of the variable Frequency, which always has a relative importance above 70% according to the profile method, and above 55% according to the perturb method, or the variable Cleanliness, whose importance is always under 50% according to the profile method, and under 20% according to the perturb method. The results of the other two methods, specially the partial derivatives method, show that any variable can reach any relative importance value between 0% and 100%. In either case, the ranking of importance of these variables is too unpredictable for the four methods. This is a major limitation for adequately determining the relative importance of the variables. However, if the values obtained individually for every variable are averaged for every method, results are more homogeneous.

(Figure 2 here)

(Figure 3 here)

(Figure 4 here)

(Figure 5 here)

In the profile method, a range of 50 profiles of variation within the interval [0,1] was generated (Figure 6). The relative importance of a variable is given by the difference between the maximum and minimum values (difference in the axis of ordinates) of the line representing the average of the profile of variation. Table 2 lists the results obtained, which indicate that the variable Frequency reaches the highest relative importance (100%), followed by Speed (77.72%), Information (64.15%) and Proximity (60.24%); so these variables have a very high global importance. A second level of importance, considered as high, includes the variables Punctuality (54.45%), Safety (53.28%) and Courtesy (48.59%), followed by a third medium-importance level containing the variables Temperature (38.44%), Fare (36.40%) and Space (27.22%). In the last place of relative importance are the variables Accesibility (17.34%) and Cleanliness (3.36%).

(Figure 6 here)

A range of 50 profiles for every variable was also generated in the perturb method, representing the profiles of variation of the MSE error, as a function of the noise percentage or perturbation introduced (Figure 7). Once again, the difference between the maximum and minimum values of the line representing the average of the profile values indicates the relative importance of the variables. The higher the variation of the MSE, the higher is the relative importance of a variable. Table 2 shows the results obtained after applying this method. Frequency (100%) is globally the most important variable, followed by Speed (63.60%). The variables Information (42.88%), Punctuality (32.92%), Safety (32.92%), Courtesy (30.33%), Proximity (23.28%) and Temperature (22.27%) belong to a second medium importance level. In the last places of the ranking appear the variables Fare (17.90%), Space (14.51%), Accesibility (7.58%) and Cleanliness (7.12%).

(Figure 7 here)

(Table 2 here)

Regarding the connection weights and partial derivatives methods, the corresponding formulas are applied to determine the relative importance of every variable (Equations 1 and 9, respectively), and the average values and ranking have been obtained and compiled in Table 2.

The connection weights method shows that the variable with the highest importance is Frequency (100%), followed by Speed (75.98%) and Information (66.68%). In a second level of high importance are included the variables Proximity (55.49%), Safety (51.38%), Punctuality (51.35%) and Courtesy (47.81%), followed by the group of medium importance variables containing Temperature (36.65%), Space (36.45%), Fare (31.98%) and Cleanliness (27.39%). The variable with less relative importance is Accessibility (14.56%).

In the partial derivatives method, the most important variables are Frequency (100%), Speed (72.90%), Information (71.01%), Punctuality (64.13%), Courtesy (60.67%) and Proximity (58.30%). The relative importance of the remaining variables can be considered as medium: Temperature (49.05%), Safety (48.08%), Fare (45.43%), Cleanliness (44.01%), Space (43.35%) and Accessibility (37.41%).

Figure 8 shows a comparison of relative importance ranking of the 12 independent variables determined by every method. This figure shows that the relative importance are very similar, and therefore the high instability inherent in these methods when applied to a single NN is considerably eliminated. The four methods agree on the fact that the most influencing variables are Frequency, Speed and Information, and that the least influencing variables are Cleanliness, Accessibility and Space, with the sole exception of the connection weights method, which considers the variables Accessibility, Cleanliness and Fare to be the least important, followed by the variable Space. The six remaining variables present intermediate positions in the ranking. The profile and perturb methods provide a more similar hierarchy of importance, while the partial derivatives method presents more discrepancies.

(Figure 8 here)

With regards to the degree of relative importance assigned to every variable, expressed as a relative percentage, it is noticed that in the profile and perturb methods, the differences allow the variables to be classified into four levels of importance: very high, high, medium and low. The perturb method allows a clear distinction to be made between three levels of importance: high, medium and low. The partial derivatives method allocates percentages of influence that can be differentiated into two levels of importance: high and medium, since the lowest value of importance is above 37%.

4. DISCUSSION

This paper, as demonstrated by many other previous studies (Mussone et al., 1999; Delen et al., 2006; Mohammadipour and Alavi, 2009; Moghaddam et al., 2010; Akin y Akbaç, 2010), confirms the high prediction capacity in the generalization stage of NN, attaining MAPE´s values below 0.05, which means more than 95% of right answers in most of the trained neural networks.

Several authors (Gedeon, 1997; Sung 1998; Palmer and Montaño, 2002b; Olden and Jackson, 2002; Gevrey et al., 2003; Paliwal and Kumar, 2.011) have analyzed the advantages and disadvantages of the existing methods to determine the relative contribution of the variables in the NN models by comparing them, but it has not been possible to establish a global consensus on which method is the most stable, accurate or robust. This leaves it open to doubt whether these methods, which determine the cause-effect relations between the predictor and dependent variables, are indeed clarifying approaches that explain the role of every variable in NN models, the so-called black boxes, and most importantly, if they are reliable enough to be used in other research fields where their results can have serious and compromising consequences.

The values of relative importance obtained by applying the profile, perturb, connection weights and partial derivatives methods to a single NN show high variability, not only when different methods are applied to the same NN, but also when one of them is applied several times to a certain NN architecture that has been trained with different initial random weights. Therefore, working this way does not guarantee the validity of the results of relative importance. However, in this study we worked with NN sets of the same architecture on which the above methods were applied. This approach achieved a similar ranking of relative importance regardless of the method used, and the results are particularly similar for the profile and perturb methods. Thus, the main goal was attained.

Any method can be considered valid, since all of them clearly agree to identify the most and least important variables, although the partial derivatives method is the least recommended because it shows a higher variability of the relative importance values.

In addition, another argument that supports the robustness of this new approach is the fact that the variables Frequency, Speed, Punctuality and Proximity are classified as the most important by all methods, and this classification agrees with the results obtained by authors (Eboli and Mazulla, 2008; Ebolli and Mazulla 2010; Dell´Olio et al.; 2010; Dell´Olio et al.; 2011; De Oña et al., 2012) who have used other techniques, such as multinomial logit models, multinomial discrete choice models, ordered probit models and decision trees.

The main weakness of this approach is that, even though the four methods provide similar results in terms of ranking variable importance, the percentages of relative importance are significantly different depending on the method used, and therefore the opinion of an expert is necessary to decide which method offers the most similar results to those expected. Thus, while the profile and perturb methods deliver a wide range of values within the interval [3.36;100.0] (the former) and [7.1;100.0] (the latter), the connection weights method limits this range to the interval [14.5;100.0] and the partial derivatives method to the interval [37.4;100.0]. The importance assigned to intermediate-positioned variables also differs from method to method.

5. CONCLUSIONS

This paper presents an approach that mitigates the instability problems inherent in the methods used to determine the contribution of predictor variables in a NN model. The principle of this approach is based, not on the modification of existing methods or the introduction of new ones, but on the application of these methods to a set of NNs instead of a single NN.

A set of ANNs with the same architecture was selected on which several existing and previously used methods were applied to determine the contribution of variables. Afterwards, a new treatment of the results was carried out, based on the calculation of the average values of the relative importance of variables.

The database used comes from a survey conducted on users of the bus transit service to know the service quality as perceived by them in the Metropolitan Area of Granada. This survey was carried out for different purposes to those pursued in this paper, and in addition it has been analyzed with different techniques in other studies (De Oña et al., 2012; 2013). We would like to point out that, to the authors' knowledge, this is the first time that ANNs are used to analyze service quality.

This approach seems to be stable, since all the methods used assign a very similar hierarchy of importance to the variables, especially to those that are have a greater impact. Additionally, these results concur with those from other studies related to the subject, which support the validity of the results.

There are, however, some differences in the importance percentage assigned to the variables depending on the method applied, since there is no single unique subset of the most important variables because of the inter-correlation of variables, and this importance is limited by the individual metric. Therefore the opinion of an expert is necessary to evaluate which method shows importance values more concordant with the expected results.

The advantages of this approach overcome the drawbacks, since it is achieved the goal of significantly eliminating the high instability existing in current methods, as they are applied so far in the NN field.

This new perspective on the application of classical methods to NN sets offers great possibilities for future research, which could study what happens with other existing methods for determining the contribution of variables, such as the recent DSA method (Cortez and Embrechts, 2013), with other preprocessing approaches for standardizing the input values, or with other typologies and architectures of NNs, with the aim of checking the robustness showed by this approach.

In terms of future work, it would be interesting to measure, after identifying most relevant inputs by methods, how a particular input tends to affect the NN output response. This is valuable in real-world applications and, for instance, it has been applied in Cortez and Embrechts (2013) using VEC curves, which are graphical representations that visualize these changes in the NN output response.

ACKNOWLEDGEMENTS

Support from Consejería de Innovación, Ciencia y Economía of the Junta de Andalucía (Spain) (Research Project P08-TEP-03819, co-funded by FEDER) is gratefully acknowledged. The authors also acknowledge the Granada Consorcio de Transportes for making the data set available for this study.

REFERENCES

Akin, D. & Akbaç, B. (2010). A neural network (NN) model to predict intersection crashes based upon driver, vehicle and roadway surface characteristics. *Scientific Research and Essays, 5(19), 2837-2847*.

Azadeh, A., Rouzbahman, M., Saberi, M. & Fam, I. M., (2011). An adaptative neural network algoritm for assessment and improvement of job satisfaction with respect to HSE and ergonomics program: the case of a gas refinery. *Journal of Loss Prevention in the Process Industries, 24, 361-370*.

Beale, M.H., Hagan, M.T. & Demuth, H.B., (2007). Neural Network Toolbox 7. User´s Guide. *MathWorks, Inc. 3 Apple Hill Drive Natic, MA 01760-2098*.

Cao, M. & Qiao, P. (2008). Neural network committee-based sensitivity analysis strategy for geotechnical engineering problems. *Neural Computing and Applications, 17, 509-519*.

Cortez, P. & Embrechts, M.J., (2011). Opening black box data mining models using sensitivity analysis. *IEEE Synopsium Series in Computational Intelligence, Paris, France, 4, 2011*.

Cortez, P. & Embrechts, M.J., (2013). Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences, 225, 1-17*.

De Oña, J., De Oña, R. & Calvo, F.J., (2012). A classification tree approach to identify key factors of transit service quality. *Expert Systems with Applications, 39, 11164-11171*.

De Oña, J., De Oña, R., Eboli, L. & Mazzulla, G., (2013). Perceived service quality in bus transit service: A structural equation approach. *Transport Policy, 29, 219-226.*

Delen, D., Sharda, R. & Bessonov, M., (2006). Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks*. Accident Analysis and Prevention*, 38, 434-444.

Dell´Olio, L., Ibeas, A. & Cecín, P. (2010). Modelling user perception of bus transit quality. *Transport Policy, 17(6), 388-397.*

Dell´Olio, L., Ibeas, A. & Cecín, P. (2011). The quality of service desired by public transport users. *Transport Policy, 18(1), 217-227*.

Dimopoulos, Y., Bourret, P. & Lek, S., (1995). Use of some sensitivity criteria for choosing networks with good generalization ability. *Neural Processing Letters, 2, 1-4*.

Eboli, L. & Mazulla, G., (2008). Willingness-to-pay of public transport users for improvement in service quality. *European Transport, 38, 107-118*.

Eboli, L. & Mazulla, G., (2010). How to capture the passengers´point of view on a transit service through rating and choice opinions. *Transport Review, 30, 435-450*.

Engelbrecht, A.P., Cloete, I. & Zurada, J.M., (1995). Determining the significance of input parameters using sensitivity analysis, from natural to artificial neural computation. *Proceedings of International Workshop on Artificial Neural Networks, pp. 382-388. Málaga-Torremolinos, Spain, Springer*.

Funahashi, K.I., (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2, 183-192.

Garson, G.D., (1991). Interpreting neural-network connection weights. *Artificial Intelligence Expert, 6, 47-51*.

Gedeon, T.D., Wong, P.M. & Harris, D., (1995). Balancing the bias and variance: network topology and pattern set reduction techniques. *Proceedings of International Workshop on Artificial Neural Networks, IWANN95, 550-558, Torremolinos, España*.

Gedeon, T.D., (1997). Data mining of inputs: analyzing magnitude of functional measures. *International Journal of Neural Systems, 8(2), 209-218.*

Gevrey, M., Dimopoulos, I. & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling, 160, 249-264*.

Gevrey, M., Dimopoulos, I. & Lek, S. (2006). Two-way interaction of input variables in the sensitivity analysis of neural network models. *Ecological Modelling, 195, 43-50*.

Goh, A.T.C., (1995). Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering, 9, 143-151*.

Hagan, M.T., Demuth, H.B. & Beale, M.H., (1996). Neural network design. *Campus Publishing Service. Colorado University Bookstore. ISBN 0-9717321-0-8*.

He, F., Sung, A.H. & Guo, B. (1997). A neural network for prediction of oil well cement bonding quality. *Proceedings of IASTED International Conference on Control, 417-420. Cancun-Mexico: IASTED-ACTA Press*.

Hunter, A., Kennedy, L., Henry, J. & Ferguson, I., (2000). Application of neural networks and sensitivity analysis to improved prediction of trauma survival. *Computer Methods and Programs in Biomedicine, 62, 11-19*.

Hornik, K., Stichcombe, M. & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks, 2, 359-366*.

Kemp, S.L., Zaradic, P. & Hansen, F., (2007). An approach for determining relative input parameter importance and significance in artificial neural networks. *Ecological Modelling. 204, 326-334*.

Kewley, R., Embrechts, M. & Breneman, C. (2000). Data strip mining for the virtual design of pharmaceuticals with neural networks. *IEEE Transactions on Neural Networks, 11(3), 668-679*.

Lek, S., Beland, A., Dimopoulos, I., Lauga, J. & Moreau, J. (1995). Improved estimation, using neural networks, of the food consumption of fish populations. *Marine and Freshwater Research, 46(8), 1229-1236*.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., & Aulagnier, S. (1996a). Application of neural networks to modeling nonlinear relationships in ecology. *Ecological Modelling, 90, 39-52*.

Lek, S., Beland, A., Baran, P., Dimopoulos, I. & Delacoste, M., (1996b). Role of some environmental variables in trout abundance models using neural networks. *Aquatic Living Resources, 9, 23-29*.

Lin, Y. & Cunningham, G.A. (1995). A new approach to fuzzy-neural system modeling. *IEEE Transactions on Fuzzy Systems, 3(2), 190-198*.

Martín del Bío, B. & Sanz Molina, A., (2.006). Neural networks and fuzzy systems. *Editorial RA-MA*.

Masters, T. (1993). Practical neural networks recipes in C++. Academic Press.

Moghaddam, F.R., Afandizadeh, S. & Ziyadi, M., (2010). Prediction of accident severity using artificial neural networks. *Internativonal Journal of Civil Ingeniering, vol 9, nº 1*.

Mohammadipour, A.H. & Alavi, S.H., (2009). The optimization of the geometric cross-section dimensions of raised pedestrian crosswalks: a case of study in Qazvin. *Accident Analysis and Prevention, 41, 314-326*.

Mussone, L., Ferrari, A. & Oneta, M., (1999). An analysis of urban collisions using an artificial intelligence model. *Accident Analysis and Prevention, 31, 705-718*.

Olden, J.D. & 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. & 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*.

Özesmi, S.L., & Özesmi, U., (1999). An artificial neural network approach to spatial habitat modeling with interspecific interaction. *Ecological Modelling, 116, 15-31*.

Paliwal, M. & Kumar, U.A., (2011). Assessing the contribution of variables in feed forward neural network*. Applied Soft Computing, 3690-3696*.

Palmer, A. & Montaño, J.J., (2002a). Redes neuronales artificiales aplicadas al análisis de datos. *Doctoral Dissertation. University of Palma de Mallorca*.

Palmer, A. & Montaño, J.J., (2002b). Redes neuronales artificiales: abriendo la caja negra. *Metodología de las ciencias del comportamiento, 4(1), 77-93*.

Palmer, A. & Montaño, J.J., (2003). Numeric Sensitivity Analysis applied to feedforward neural networks. *Neural Computing and Applications, 12, 119-125*.

Rzempoluk, E.J., (1998). Neural network data analysis using Simulnet. *New York: Springer-Verlag.*

Rumelhart, D.E. & McClelland, J.L. (1986). Parallel Distributed Processing. Vol 1: Foundations. *MIT Press*.

Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986). Learning representations by backpropagation errors. *Nature, 323, 533-536*.

Scardi, M. & Harding, L.W., (1999). Developing an empirical model of phytoplankton primary production: a neural networks case study. *Ecological Modelling, 120(2-3), 213-223*.

Sung, A.H., (1998). Ranking importance of input parameters of neural networks. *Expert Systems with Applications, 15, 405-411*.

Watts, M.J., & Worner, S.P., (2008). Using artificial neural networks to determine the relative contribution of abiotic factors influencing the establishment of insect pest species. *Ecological Informatics, 3, 64-74*.

Werbos, P.J., (1974). Beyond Regression: new tools for prediction and analysis in behavioral sciences. *Doctoral Dissertation. Applied Mathematics, Harvard University*.

Yao, J., Teng, N., Poh, H.L. & Tan, C.L., (1998). Forecasting and analysis of marketing data using neural networks. *Journal of Information Science and Engineering, 14, 843-862.*

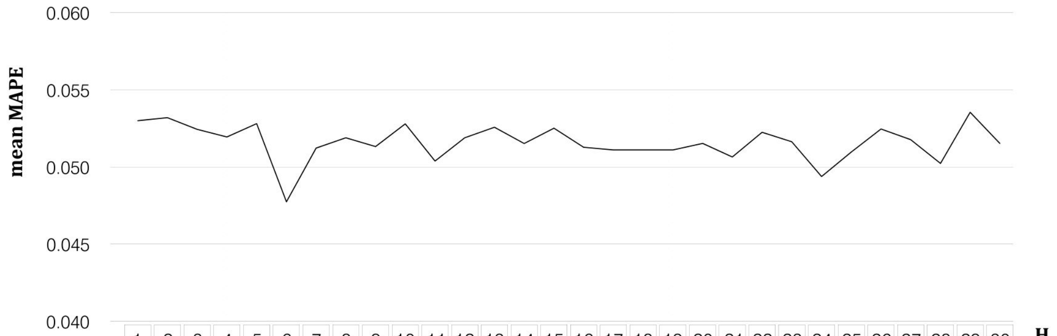
Yeh, I. & Cheng, W., (2010). First and second order sensitivity analysis of MLP. *Neurocomputing, 73, 2225-2233*.

Zhou, Z.H., Wu, J. & Tang, W., (2002). Ensembling neural networks: many could be better than all. *Artificial Intelligence 137 (1-2), 239-263*.

Zurada, J.M., Malinowski, A. & Cloete, I. (1994). Sensitivity analysis for minimization of input data dimensión for feed forward neural network. *Proceedings of IEEE International Synopsium on Circuits and Systems. London. IEEE Press*.

List of Figures:

List of Tables:



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	H
mean MAPE	0.0531	0.0532	0.0525	0.0519	0.0528	0.0475	0.0514	0.0520	0.0513	0.0530	0.0505	0.0520	0.0526	0.0517	0.0526	0.0514	0.0513	0.0513	0.0513	0.0516	0.0508	0.0524	0.0518	0.0497	0.0508	0.0524	0.0520	0.0507	0.0534	0.0517	
standard deviation	0.0082	0.0090	0.0045	0.0067	0.0062	0.0064	0.0052	0.0044	0.0036	0.0074	0.0051	0.0086	0.0106	0.0043	0.0065	0.0087	0.0056	0.0070	0.0046	0.0049	0.0056	0.0078	0.0070	0.0047	0.0067	0.0063	0.0078	0.0056	0.0075	0.0091	

Figure 1. Mean MAPE and standard deviation´s values depending on H NN subsets

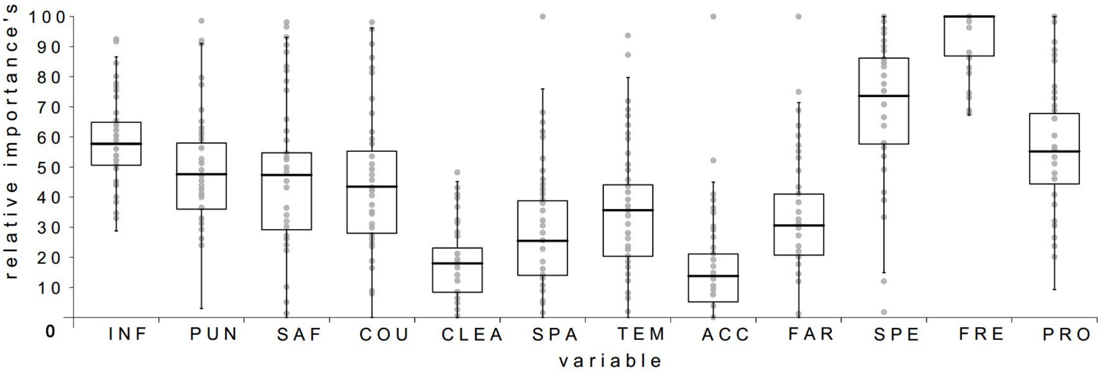


Figure 2. Boxplot of variables´ relative importance by Profile Method

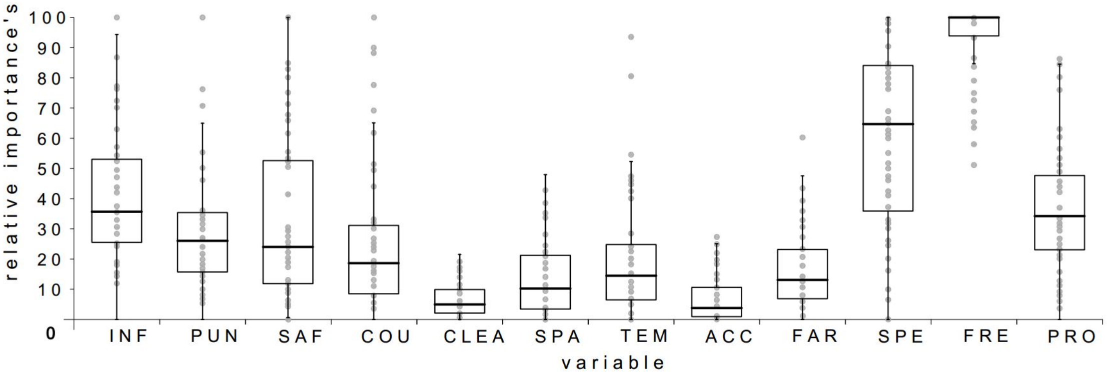


Figure 3. Boxplot of variables´ relative importance by Perturb Method

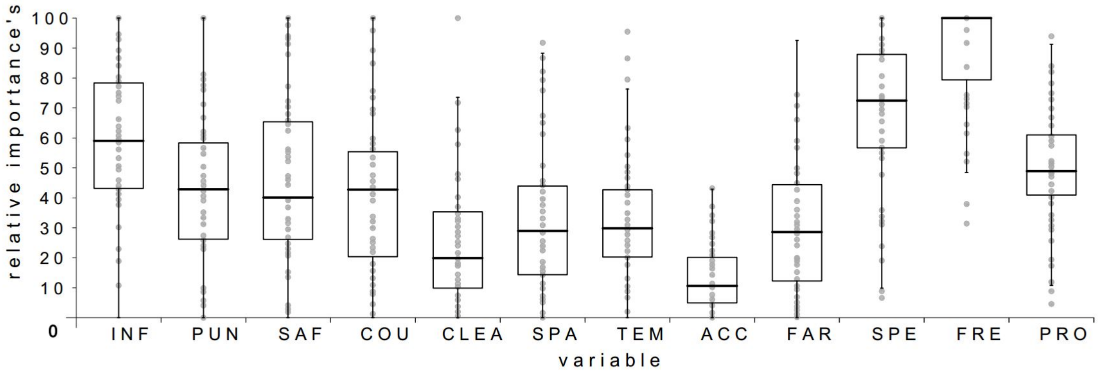


Figure 4. Boxplot of variables´ relative importance by Connection Weights Method

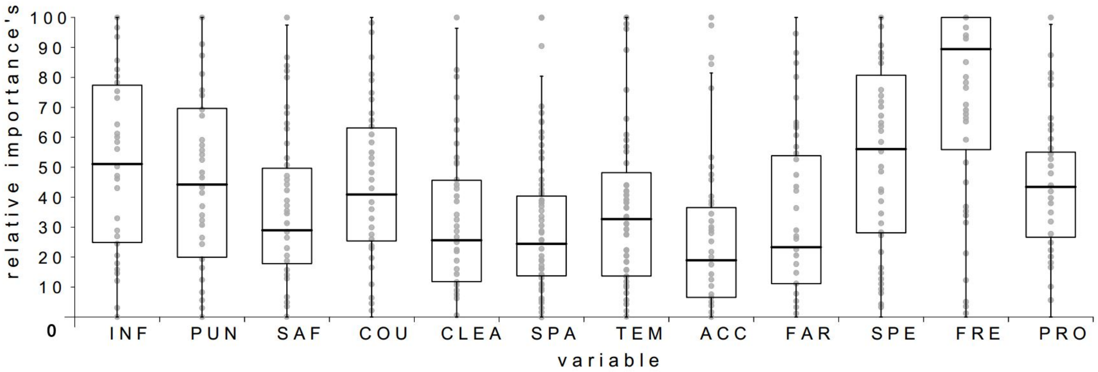


Figure 5. Boxplot of variables´ relative importance by Partial Derivates Method

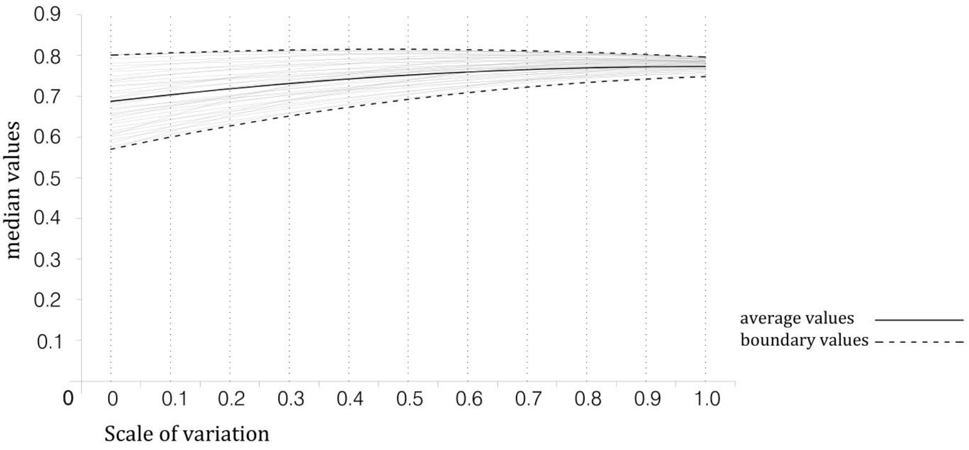


Figure 6. Average and boundary values of the variable INFORMATION Profile method

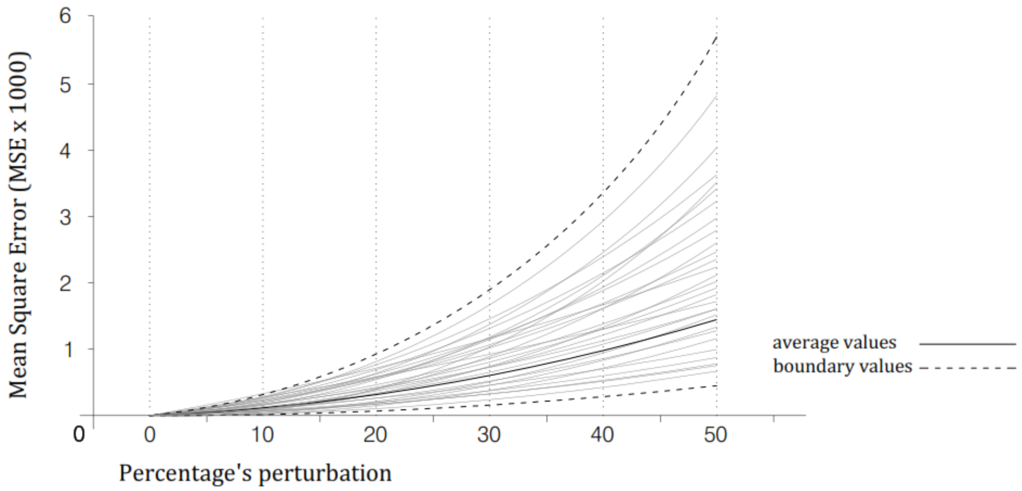


Figure 7. Average and boundary values of MSE of the variable INFORMATION Perturb method

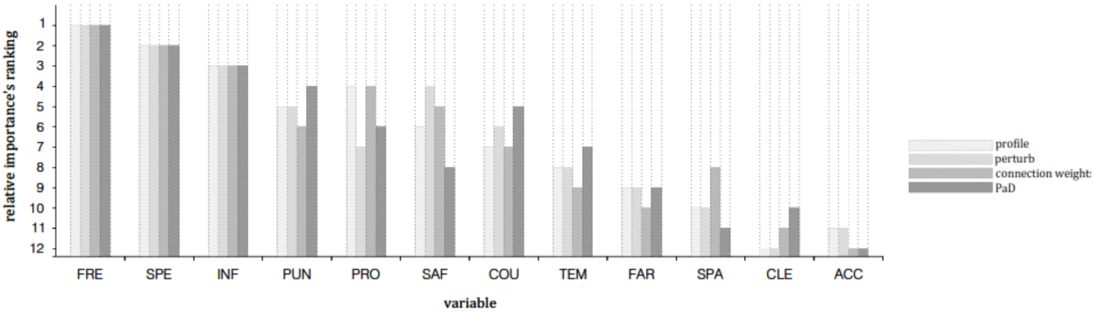


Figure 8. Comparative of ranking´s relative importance by methods.

	VARIABLE	SYMBOL	VALUES
IMPUT LAYER	**INFORMATION**	INF	[0;10]
	PUNCTUALITY	PUN	[0;10]
	SAFETY	SAF	[0;10]
	COURTESY	COU	[0;10]
	CLEANLINESS	CLE	[0;10]
	SPACE	SPA	[0;10]
	TEMPERATURE	TEM	[0;10]
	ACCESSIBILITY	ACC	[0;10]
	FARE	FAR	[0;10]
	SPEED	SPE	[0;10]
	FREQUENCY	FRE	[0;10]
	PROXIMITY	PRO	[0;10]
OUTPUT LAYER	**QUALITY OF SERVICE**	QS	[0;10]

Table 1. Customer Satisfaction Survey's Items

VARIABLE	DERIVED IMPORTANCE							
	PROFILE		PERTURB		CONNECTION WEIGHTS		PARTIAL DERIVATES	
	Average	Ranking	Average	Ranking	Average	Ranking	Average	Ranking
INF	64.15	3	42.88	3	66.68	3	71.01	3
PUN	54.45	5	32.92	4-5	51.35	6	64.13	4
SAF	53.28	6	32.92	4-5	51.38	5	48.08	8
COU	48.59	7	30.33	6	47.81	7	60.67	5
CLE	3.36	12	7.12	12	27.39	11	44.01	10
SPA	27.22	10	14.51	10	36.45	9	43.35	11
TEM	38.44	8	22.27	8	36.65	8	49.05	7
ACC	17.34	11	7.58	11	14.56	12	37.41	12
FAR	36.40	9	17.90	9	31.98	10	45.43	9
SPE	77.72	2	63.60	2	75.98	2	72.90	2
FRE	100.00	1	100.00	1	100.00	1	100.00	1
PRO	60.24	4	23.28	7	55.49	4	58.30	6

Table 2. Variables' relative importance by methods