A hybrid system of neural networks and rough sets for road safety performance indicators

Peer-reviewed author version

A hybrid system of neural networks and rough sets for road safety performance indicators

Yongjun Shen[1], Tianrui Li[2], Elke Hermans[1], Da Ruan[1, 3], Geert Wets[1], Koen Vanhoof[1] and Tom Brijs[1]

[1]*Transportation Research Institute – Hasselt University*
Wetenschapspark 1 bus 12, 3590 Diepenbeek, Belgium
{yongjun.shen, elke.hermans, da.ruan, geert.wets, koen.vanhoof, tom.brijs}@uhasselt.be
[2]*School of Information Science and Technology, Southwest Jiaotong University*
Chengdu 610031, China
trli@swjtu.edu.cn
[3]*Belgian Nuclear Research Centre (SCK·CEN)*
Boeretang 200, 2400 Mol, Belgium
druan@sckcen.be

Abstract
Road safety performance indicators are comprehensible tools that provide a better understanding of current safety conditions and can be used to monitor the effect of policy interventions. New insights can be gained in case one road safety index is composed of all risk indicators. The overall safety performance can then be evaluated, and countries ranked. In this paper, a promising structure of neural networks based on decision rules generated by rough sets—is proposed to develop an overall road safety index. This novel hybrid system integrates the ability of neural networks on self-learning and that of rough sets on automatically transforming data into knowledge. By means of simulation, optimal weights are assigned to seven road safety performance indicators. The ranking of 21 European countries in terms of their road safety index scores is compared to a ranking based on the number of road fatalities per million inhabitants. Evaluation results imply the feasibility of this intelligent decision support system and valuable predictive power for the road safety indicators context.

Keywords: *Road safety performance indicators; (Composite) index; Road safety output (RSO); Neural networks; Rough sets; Hybrid system; Decision support*

1 Introduction

Steadily increased traffic volume resulted in continuously increasing safety problems during the past decades. Every year, about 40,000 people die in Europe as a consequence of road crashes (European Transport Safety Council 2008). Thereby, the European Union has set itself a target of halving the yearly number of road fatalities between 2001 and 2010 (European Commission 2001). To reach this target, besides analyzing the number of accidents and casualties, road safety policymakers and analysts also need to take into account as many underlying risk factors influencing safety as possible or, at least, those factors that they are able to affect or control. At the same time, each risk factor should be represented by a set of quantifiable road safety performance indicators, which provide a means for monitoring the effectiveness of the safety actions that are taken (European Transport Safety Council 2001). Moreover, to obtain an overall idea of road safety performance, the development of a composite indicator or index, which is a combination of individual indicators in several risk domains, is helpful and desirable (Hermans et al. 2008a, b; Wegman et al. 2008). The index

can be used for reducing the complexity of the road safety problem, benchmarking and prioritizing road safety actions (Hermans et al. 2008a, b). As concluded in Nardo et al. (2005), the application of this kind of indicator is a clear symptom of its political importance and operational relevance in decision making. However, the construction of a composite indicator involves several methodological stages: the selection of indicators, data normalization, imputation of missing values, the choice of aggregation models, the weights of the indicators, and so on (Hermans 2009). Although other factors may have a certain impact on the final scores and rankings, the issue of weighting plays a central role to the development of a composite indicator. Ideally, the weighting, i.e., assessing a relative importance to the underlying individual indicators, should be made explicitly and accepted by an as wide as possible public. However, this is practically impossible, as among several methods in the composite indicator literature (Hermans et al. 2008a, b; Nardo et al. 2005) for weighting the different indicators, there is no best one to use in all circumstances.

Neural networks (NNs) and rough sets are two valuable soft computing techniques that have been widely studied and successfully applied to a variety of activities (Hornik et al. 1989; Xiang et al. 2005; Skowron 1995; Peters and Skowron 2002). Further exploration of their combination is also conducted, in which rough sets are usually utilized to reduce the set of attributes followed by NNs (Luo and Shao 2006; Jiang et al. 2006; Xiao et al. 2009). In this paper, a new idea of the integration of these two techniques is proposed and applied as a novel weighting method. More specifically, we integrate these two techniques into an intelligent decision support system and cooperatively obtain weights for the road safety performance indicators and subsequently composite index values. In this hybrid framework, two neural networks—a single neuron and a two-layer feed-forward network called a gating network—are introduced to calculate the weight of each indicator by self-learning based on their remarkable ability to derive meaning from complicated or imprecise data that are too complex to be noticed by either humans or other computer techniques. At the same time, in order to deal with two main uncertain factors linked to neural networks training, i.e., the initial value of network weights and the number of iterations, rough sets theory, which is capable of rule induction and feature selection in decision situations, is utilized to determine the most important road safety attribute and therefore generating a decision rule to guide the network training.

The remaining of the paper is organized as follows: in Sect. 2, the data used in this study are discussed. The construction and training algorithm of neural networks, as well as the selection of the initial value and the number of iterations are specified in Sect. 3, while Sect. 4 introduces the fundamentals of rough sets theory and its application in this study. The hybrid system based on neural networks and rough sets is illustrated in Sect. 5 and the simulation results are discussed in Sect. 6. The paper ends with conclusions and topics for further research.

2 Road safety data

Road safety performance indicators are measures that are causally related to accidents or injuries and are used in addition to the number of accidents or injuries, in order to indicate safety performance or better understand the processes that lead to accidents (European Transport Safety Council 2001). In the European SafetyNet project work package on safety performance indicators (SafetyNet Work Package 3 2005), the following seven risk domains are designated as central to road safety activities in Europe and are selected for the development of safety performance indicators: alcohol and drugs, speed, protective systems, daytime running lights, vehicle, roads, and trauma management. Moreover, each risk domain needs to be represented by at least one quantifiable performance indicator which is policy relevant and has data available and reliable (Litman 2007).

In this paper, an overall index will be constructed based on seven indicators. For alcohol and drugs, the percentage of drivers respecting the blood alcohol content (BAC) limit is the indicator (I1), the speed indicator is the percentage of drivers with a driving speed below the maximum speed limit (I2), the protective systems domain is represented by the seat belt wearing rate in the front seats (I3), the existence of a law obligating daytime running lights (DRL) is selected for the DRL domain (I4), the share of cars of maximum five years old is the selected vehicle indicator (I5), the network density describes the roads domain (I6) and for trauma management (TM) the health expenditure as share of the gross domestic product is the selected indicator (I7).

From international data sources (amongst others [SARTRE 3 Report 2004]) values related to 2003 were obtained for these seven risk indicators for 21 European countries being Austria (AT), Belgium (BE), Cyprus (CY), Czech Republic (CZ), Denmark (DK), Estonia (EE), Finland (FI), France (FR), Germany (DE), Greece (EL), Hungary (HU), Ireland (IR), Italy (IT), the Netherlands (NL), Poland (PL), Portugal (PT), Slovenia (SL), Spain (ES), Sweden (SE), Switzerland (CH), and United Kingdom (UK). To eliminate the effect of the measurement unit and the scale of each indicator, we normalize all indicator values into an interval of [0,1] with a higher value representing a better performance, as presented in Table 1. The last column corresponds to a common road safety output (RSO), i.e., the number of road fatalities per million inhabitants. Here, the values represent one minus the normalized number of road fatalities per million inhabitants with a high score also implying a good performance. In other words, Sweden has the lowest number of road fatalities per million inhabitants and is considered the safest country (in 2003 fatality terms). The countries' ranking based on the constructed index consisting of seven road safety performance indicators should approximate the RSO ranking to the utmost extent. In the following sections, the methodology is described in detail.

Table 1. Data on the seven road safety risk indicators and fatalities

	Alc.	Speed	Prot. Syst.	DRL	Vehicle	Roads	TM	Outcome
	I_1	I_2	I_3	I_4	I_5	I_6	I_7	
AT	0.88	0.56	0.65	1.00	0.48	0.29	0.43	0.38
BE	0.73	0.38	0.46	0.50	0.64	1.00	0.66	0.36
CY	0.00	0.13	0.70	0.00	0.00	0.22	0.21	0.17
CZ	0.91	0.81	0.37	0.50	0.06	0.30	0.41	0.08
DK	0.99	0.38	0.77	1.00	0.59	0.31	0.61	0.77
EE	0.93	0.00	0.61	1.00	0.00	0.22	0.00	0.31
FI	0.99	0.63	0.86	1.00	0.32	0.00	0.36	0.84
FR	0.77	0.63	1.00	0.50	0.54	0.34	0.75	0.53
DE	0.89	0.19	0.95	0.50	0.56	0.34	0.95	0.77
EL	0.64	0.06	0.00	0.00	0.29	0.14	0.72	0.03
HU	0.94	0.19	0.33	0.50	0.25	0.32	0.44	0.20
IR	0.89	1.00	0.79	0.00	1.00	0.24	0.36	0.72
IT	0.67	0.31	0.54	0.50	0.49	0.30	0.56	0.49
NL	0.91	0.38	0.81	0.00	0.58	0.60	0.66	0.96
PL	0.99	0.56	0.54	0.50	0.14	0.24	0.16	0.00
PT	0.81	0.31	0.84	0.50	0.31	0.14	0.69	0.01
SL	0.88	0.63	0.72	1.00	0.88	0.36	0.51	0.31
ES	0.67	0.44	0.81	0.50	0.51	0.10	0.41	0.21
SE	1.00	0.38	0.91	1.00	0.60	0.05	0.67	1.00
CH	0.81	0.13	0.70	0.50	0.60	0.32	1.00	0.82
UK	0.97	0.75	0.93	0.00	0.73	0.28	0.43	0.98

3 Neural networks: theory and application

Neural networks are computational architectures that combine simple units in an arrangement that can then exhibit complex behavior (Haykin 1999). One of the most familiar architectures is the feed-forward NN (Hornik et al. 1989; Xiang et al. 2005). Over the years, NNs have proven to be a very powerful tool capable of extracting reliable information and patterns from complicated data even in the absence of models describing the data (Reich and Baraib 2000). In this section, a composite feed-forward NN will be illustrated and experimented with in order to obtain the weights of seven road safety performance indicators.
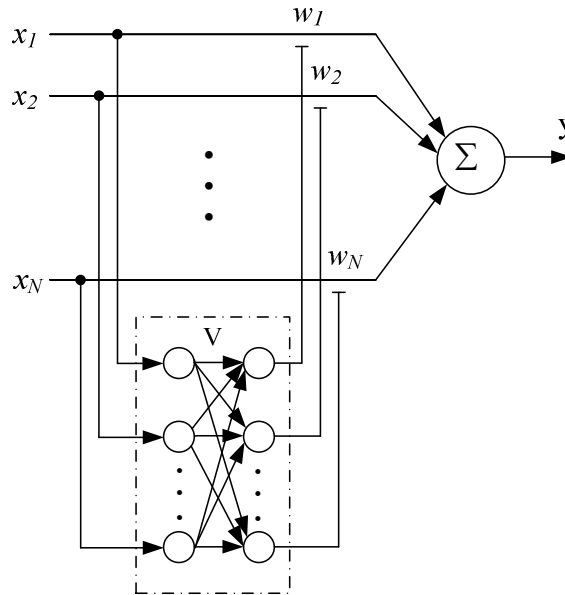
3.1 Construction of neural networks

In the case of aggregating N road safety performance indicators into one combined index, a typical composite indicator takes the following form (Nardo et al. 2005):

$$y = \sum_{i=1}^{N} w_i x_i \tag{1}$$

where y represents the composite index, x_i the ith normalized indicator and w_i the weight assigned to x_i, which should satisfy two restrictions:

$$0 \le w_i \le 1 \qquad 1 \le i \le N$$

$$\sum_{i=1}^{N} w_i = 1 \tag{2}$$

In order to characterize the above composite indicator, two neural networks are introduced to work in unison, whose structure is shown in Figure 1. One (the upper part) is a single neuron used to calculate the composite index, see (1), and the other one (the lower part) is a two-layer feed-forward network with no hidden layers in it, the so-called gating network, aimed at assigning weights to the corresponding indicators, thereby guaranteeing the satisfaction of (2) (Shen et al. 2008).

Figure 1. Construction of neural networks

In this respect, an intermediary variable ξ_i is introduced, and the ith output of the gating network is calculated as follows:

$$w_i = \frac{e^{\xi_i}}{\sum_n e^{\xi_n}} \tag{3}$$

In (3),

$$\xi_i = v_i^T x \tag{4}$$

where $v_i\ (i = 1, 2, \cdots N)$ are the vectors of the matrix V , i.e., the weight coefficients of the gating network.

3.2 Network Training

By definition of the gating network, the tuning of the weights w_i is transformed into the self-learning of the weights of the gating network, i.e., V. To guarantee the accuracy and stability of the system, the application of a learning algorithm, the selection of the initial value of V and the number of iterations are three essential aspects.

3.2.1 Learning Algorithm

In the NNs training, various error back propagation (EBP) algorithms (Rumelhart et al. 1986; Werbos 1988) exist, such as the gradient descent method, the conjugate gradient method, the Gauss–Newton algorithm, and the Levenberg–Marquardt (LM) algorithm. Different types of learning algorithms affect the speed and accuracy of the training to a great extent. Among them, the LM algorithm as a blend of the gradient descent and Gauss–Newton method is the most widely used optimization algorithm since it converges in a much faster way by using the second-order information (Hagan and Menhaj 1994). The learning process is as follows.

The output of the network has to be judged using some error criterion. This criterion determines the size of the error to be back propagated. In general, the mean squared error (MSE) criterion is used:

$$E(k) = \frac{1}{2}\frac{1}{M}\sum_{m=1}^{M}[y'(m) - y(m)]^2 = \frac{1}{2M}\sum_{m=1}^{M}e^2(m) \tag{5}$$

where y' represents the desired output value, i.e., the RSO value in this case, M is the size of the learning set, and k the number of iterations.

The objective during the training is to minimize the MSE by choosing appropriate weights, which are adjusted in the following way:

$$v^{(k+1)} = v^{(k)} - [J^T J + \mu I]^{-1} J^T e \tag{6}$$

where J is the Jacobi matrix as calculated as follows:

$$J = \begin{pmatrix} \dfrac{\partial e_1}{\partial v_1} & \dfrac{\partial e_1}{\partial v_2} & \cdots & \dfrac{\partial e_1}{\partial v_N} \\ \dfrac{\partial e_2}{\partial v_1} & \dfrac{\partial e_2}{\partial v_2} & \cdots & \dfrac{\partial e_2}{\partial v_N} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial e_M}{\partial v_1} & \dfrac{\partial e_M}{\partial v_2} & \cdots & \dfrac{\partial e_M}{\partial v_N} \end{pmatrix} \tag{7}$$

I denotes the identity matrix with the same size as J, and μ is the learning rate (e.g., 0.01). The introduction of the term μI is to guarantee the existence of the inverse of matrix $J^T J$, which is used as an approximation of the second derivative Hessian matrix.

3.2.2 Initial Value

The selection of the initial weight coefficients is one of the most fundamental and important concepts in the study of NNs because ill-conditioning is a common cause of slow and inaccurate or even divergent results from back propagation algorithms (Atiya and Ji 1997). Following the construction of NNs and the LM algorithm specified in the previous section, the NNs are applied to the data in Table 1 to construct a composite indicator starting from several sets of random initial weights.

In the simulation, the seven road safety indicators are presented as the inputs of the NN and the RSO as the system derived output. The number of iterations equals 1,000. The assigned weights are shown in Table 2. It can be seen that different initial values lead to different weights within the same algorithm and the same number of iterations. Although indicators I3 and I5 corresponding to the protective systems and vehicle domain get relatively high weights, while the share assigned to the DRL domain (I4) is the lowest in all experiments, there is no uniform result and this causes a difficulty in interpretation.

Table 2. Weights after 1000 iterations based on several random initial values

Initial value	I_1	I_2	I_3	I_4	I_5	I_6	I_7
random1	0.040	0.028	0.339	0.005	0.391	0.055	0.143
random2	0.038	0.017	0.342	0.009	0.416	0.055	0.124
random3	0.040	0.020	0.312	0.013	0.421	0.038	0.156

3.2.3 Number of Iterations

Besides the selection of the initial weights the iteration time is another uncertain factor in the self-learning of NNs which alter the final results to a certain extent. In the current case, the same learning algorithm is used, and the same initial weights are given, i.e., a zero matrix is selected as the initial value of V. After the calculation of (3) and (4), ξ_i is equal to zero and w_i results in $1/N$, which implies that at the first step of iteration, equal weighting is applied. It guarantees the initial stability of the system. Then, after 400, 800, and 1,000 iterations, respectively, different results are obtained in Table 3.

Table 3. Weights after different iteration times starting from equal weights

Iterations	I_1	I_2	I_3	I_4	I_5	I_6	I_7
400	0.068	0.059	0.294	0.025	0.318	0.076	0.160
800	0.042	0.03	0.332	0.010	0.384	0.058	0.143
1000	0.035	0.023	0.342	0.007	0.401	0.054	0.138

Table 3 shows the final weights of the corresponding road safety indicators after different numbers of iterations. Along with the increase of the number of iterations, the weights of indicators I_3 and I_5 go upward, while the shares of the other indicators decline, especially the DRL domain, which even tends to zero. However, there are no certain constraint conditions to determine the best time to stop the algorithm.

In the next section, rough sets theory as a useful tool for both rule induction and feature selection will be explained and applied to guide the network training by generating a decision rule to determine the initial weights and the iteration time.

4 Rough Sets

Rough sets theory originated by Pawlak is a formal mathematical theory modeling knowledge about the domain of interest in terms of a collection of equivalence relations (Pawlak 1991). It is mainly an automated transformation of data into knowledge without any preliminary or additional information about the data like probability in probability theory, or grade of membership in fuzzy set theory. It is based on the concept of an upper and a lower approximation of a set, approximation space, reduct and core, etc. The concepts in rough sets theory are used to define the necessity of features. The measures of necessity are calculated by the lower and upper approximations. These measures are employed as heuristics to guide the feature selection process. Nowadays, many rough sets based approaches have been successfully applied in the field of knowledge discovery (Skowron 1995; Peters and Skowron 2002; Cao et al. 2003, Pawlak and Skowron 2007).

4.1 Fundamentals of Rough Sets Theory

Basic concepts and relevant notions of rough sets theory are briefly cited from Swiniarski and Skowron (2003), Pawlak et al. (2005) and Li et al. (2007). An information system can be represented as $S = (U, A, V, f)$, where U is the *universe*, a finite set of n objects $\{x_1, x_2, \ldots, x_n\}$. A is a finite set of *attributes*, and $A = C \cup D$, where C is a set of *condition* attributes, and D is a set of *decision* attributes. $V = \bigcup_{a \in A} V_a$, where V_a is a domain of the attribute a. $f : U \times A \to V$ is a *decision function* (called the *information function*) such that $f(x, a) \in V_a$ for every $a \in A, x \in U$. A subset of attributes $P \subseteq A$ defines an *indiscernibility relation* on U:

$$IND(P) = \{(x, y) \in U \times U : f(x, a) = f(y, a), \forall a \in P\} \tag{8}$$

Obviously, the indiscernibility relation is an equivalence relation (Swiniarski 2001). $IND(P)$ is the Partition of U determined by P. The family of all equivalence classes of $IND(P)$ (Partition of U determined by P) is denoted by $U/IND(P)$. Each element in $U/IND(P)$ is a set of indiscernible objects with respect to P. Equivalence classes $U/IND(C)$ and $U/IND(D)$ are called condition and decision classes. For any concept $X \subseteq U$ and attribute subset $P \subseteq A$, X could be approximated by the *P*-lower approximation and *P*-upper approximation using the knowledge of P. The lower approximation of X is the set of objects of U that are surely in X, defined as:

$$\underline{P}(X) = \bigcup \{B \in U / IND(P) : B \subseteq X\} \tag{9}$$

The upper approximation of X is the set of objects of U that are possibly in X, defined as:

$$\overline{P}(X) = \bigcup \{B \in U / IND(P) : B \cap X \neq \varnothing\} \tag{10}$$

The boundary region is defined as:

$$BND_P(X) = \overline{P}(X) - \underline{P}(X) \tag{11}$$

If the boundary region is empty, that is, $\overline{P}(X) = \underline{P}(X)$, concept X is said to be *P*-definable. Otherwise X is a rough set with respect to P.

An attribute $a \in A$ is called dispensable in the set A if $IND(A) = IND(A-\{a\})$ (it means that indiscernibility relations generated by sets A and $A-\{a\}$ are identical). Otherwise the attribute a is indispensable in A. Obviously, the dispensable attribute does not improve the classification of the original information system. The set $P \subseteq A$ is called orthogonal if all its attributes are indispensable. A proper subset $R \subset A$ is defined as a *reduct* set of A if R is orthogonal and preserves the classification generated by A. The reduct set of A, denoted by $RED(A)$, is defined as:

$$RED(A) = \{R : R \subset A, IND(R) = IND(A), R \text{ is orthogonal}\} \tag{12}$$

A core of A is the intersection of all the reducts of A:

$$CORE(A) = \bigcap RED(A) \tag{13}$$

The core contains all the attributes that cannot be removed from the set A without changing the original classification.

4.2 Application of Rough Sets

The reduct and core are two fundamental concepts of rough sets theory that can be used for both feature selection and data reduction. A reduct is the essential part of an information system which can discern all objects by the original set of attributes of the information system. A core is a common part of all reducts. It is the most important attribute of an information system since it cannot be removed from the information system without causing collapse of the equivalence class structure. By this feature, we determine the most important attribute or indicator (with the highest weight) for the final calculation of the composite indicator.

Most of the road safety performance indicator measurements are real numbers and have to be discretized before the rough sets approach can be applied. A discretization method finds a set of cuts that divides the range of an attribute a into a set of intervals or bins. Let

$C_a = \{c_1^a, \ldots, c_i^a, \ldots, c_n^a\}$ be such a set of cuts for attribute a, where $c_1^a < \cdots < c_i^a < \cdots < c_n^a$. The bins are then defined as (Midelfart et al. 2002):

$$\begin{cases} B_0 = \{x \in U : a(x) < c_1^a\} \\ B_i = \{x \in U : c_i^a \leq a(x) < c_{i+1}^a\}, 1 \leq i < n-1 \\ B_n = \{x \in U : c_n^a \leq a(x)\} \end{cases} \tag{14}$$

Several discretization methods exist, and equal frequency binning is used in this study. This is an unsupervised method that divides the range into n intervals so that the frequency of the objects is the same in each interval, i.e., $|U_i|=|U_j|$ for all $0 \leq i$, $j \leq n-1$. It is generally better than equal interval binning since an uneven distribution might seriously impair the ability of the attribute to build good decision structures. The discretized data (here we use 1, 2, 3, and 4 to represent the bins) are shown in Table 4 by using the equal frequency binning discretization method—PKIDiscretize, in Weka (Witten and Frank 2005).

Table 4 Discretized data on the seven road safety performance indicators and the road fatalities

	Alc. (I_1)	Speed (I_2)	Prot. syst. (I_3)	DRL (I_4)	Vehicle (I_5)	Roads (I_6)	TM (I_7)	RSO
AT	2	3	2	3	2	3	2	3
BE	1	2	1	2	4	4	3	4
CY	1	1	2	1	1	2	1	4
CZ	3	4	1	2	1	3	2	2
DK	4	2	3	3	3	3	3	2
EE	3	1	2	3	1	2	1	3
FI	4	3	4	3	2	1	1	1
FR	2	3	4	2	3	4	4	1
DE	3	1	4	2	3	4	4	1
EL	1	1	1	1	2	1	4	3
HU	3	1	1	2	1	3	2	3
IR	3	4	3	1	4	2	1	3
IT	1	2	1	2	2	3	3	4
NL	3	2	3	1	3	4	3	4
PL	4	3	1	2	1	2	1	4
PT	2	2	3	2	2	1	4	2
SL	2	3	2	3	4	4	3	2
ES	1	3	3	2	3	1	2	2
SE	4	2	4	3	3	1	3	1
CH	2	1	2	2	4	3	4	1
UK	4	4	4	1	4	2	2	3

For the DRL domain, there are only three values in the raw data set, so only three discretized values are obtained

From the discretized data, all reducts could be obtained by ROSETTA (Øhrn et al. 1998):

Reducts={{Alcohol and drugs, Speed, Vehicle}; {Alcohol and drugs, Speed, Roads, Trauma Management}; {Alcohol and drugs, Speed, Protective Systems, Roads}; {Alcohol and drugs, Speed, Daytime Running Lights, Roads}; {Speed, DRL, Roads, Trauma Management}; {Speed, Protective Systems, Daytime Running Lights, Roads}; {Speed, Protective Systems, Daytime Running Lights, Vehicle}; {Speed, Daytime Running Lights, Vehicle, Trauma Management}}.

Then, calculating the intersection of these eight reducts, we get: *Core*={Speed}.

Therefore, the speed indicator appears to be the most important attribute and should be allocated the highest weight in the calculation of the composite road safety performance index. This is also validated by a recent research surveying some road safety experts (Hermans et al. 2008a, b). This result will be used to direct the training of NNs by adding the constraint condition that the speed indicator should have the highest weight.

5 Hybrid System based on Neural Networks and Rough Sets

By integrating neural networks and rough sets principles into a hybrid system, we derive a novel intelligent decision support system, i.e., a construction of neural networks based on decision rules generated by rough sets (NNRS), shown in Fig. 2.
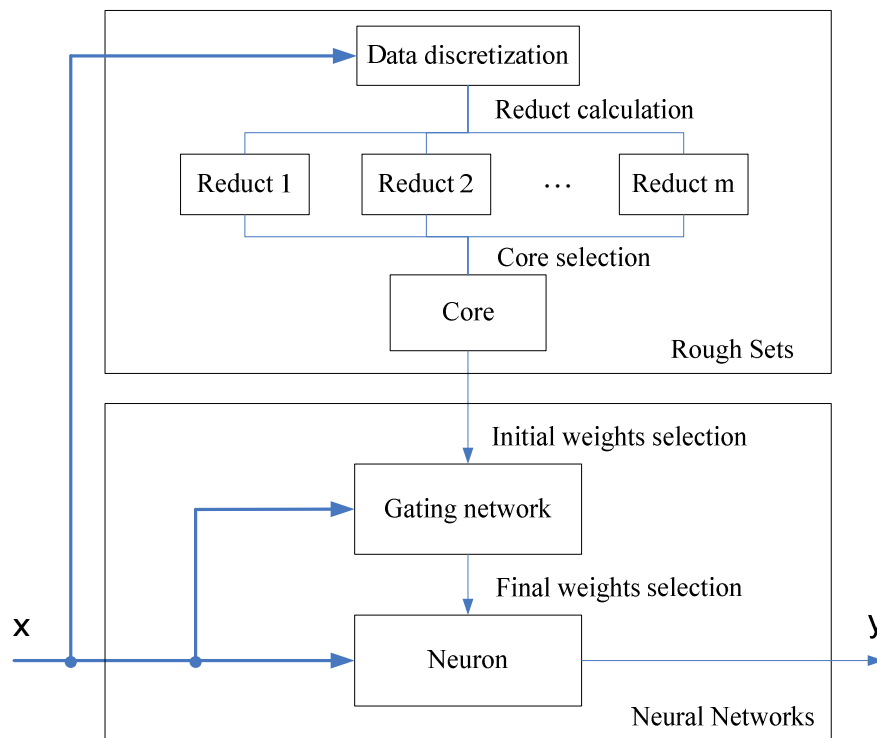
Figure 2. Construction of the hybrid system based on neural networks and rough sets

In this hybrid framework, NNs as self-learning machines are introduced to obtain the weights of the seven road safety performance indicators and subsequently the composite index scores. At the same time, in order to reduce two main uncertain elements of NNs, i.e., the initial value of the network weights and the number of iterations, rough sets theory are utilized to determine the most important attribute in this system, which is translated into a decision rule to guide the network training. The detailed steps of the methodological process are as follows, and the results are presented in the next section.

Step 1. Data discretization based on the method of equal frequency binning.
Step 2. The rough sets process:
a. Calculate the reducts of the discretized information system by ROSETTA.

b. Select the core indicators that are the intersections of all the reducts.

c. Generate the decision rules to guide the network training.

Step 3. The neural network process:

a. Assign initial weights to the gating network.

b. Network training based on the LM algorithm.

c. Obtain the optimal weights and compute the final composite indicator scores.

6 Simulation results and discussions

To guarantee the satisfaction of the constraint condition derived from the rough sets theory—the speed indicator should be allocated the highest weight—the initial values of the weights of the gating network are as follows:

$$V^{(1)} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 1 & 0 & \cdots & 0 \end{bmatrix}_{7 \times 7}$$

(15)

In the computation of the weights w_i $(i = 1, 2, \cdots, 7)$, the weight of the speed domain, i.e., w_2 gets the highest value at the beginning of the learning, then the NNRS searches for the most optimal weights resulting in a minimal MSE while respecting the 'speed' constraint. The learning algorithm is the same as before (see Section 3). The self-tuning of the neural networks stops at 572 iterations and the final weights are shown in Table 4.

In the computation of the weights w_i $(i = 1, 2, \cdots, 7)$, the weight of the speed domain, i.e., w_2 gets the highest value at the beginning of the learning. Next, the NNRS searches for the most optimal weights resulting in a minimal MSE while respecting the 'speed' constraint. The learning algorithm is the same as before (see Sect. 3). The self-tuning of the neural networks stops after 572 iterations and the final weights are shown in Table 5.

Table 5. Weights based on the hybrid system

	I_1	I_2	I_3	I_4	I_5	I_6	I_7
Weights based on NNRS	0.052	0.238	0.237	0.033	0.204	0.049	0.187

It shows that indicator I2 (corresponding to the speed domain) keeps the highest weight, followed by I3 and I5, while the weight assigned to the DRL indicator is still the lowest. This set of weights as well as the weights based on the neural networks shown in Tables 2 and 3 are respectively used to calculate the composite road safety performance index score for the 21 countries. We compute the sum of the product of each normalized indicator value and its assigned weight to obtain these scores and subsequently rank the countries in such a manner that a higher score or performance relates to a better rank. The ranking results even as the RSO ranking are shown in Table 6, where E1 to E6 represent the experiments carried out with respect to several sets of random initial weights and different numbers of iterations (see Sect. 3).

Table 6. Rankings based on NNs, NNRS and RSO

Countries	E1 ranks	E2 ranks	E3 ranks	E4 ranks	E5 ranks	E6 ranks	NNRS ranks	RSO ranks
	random1	random2	random3	400	800	1000		
AT	13	13	13	12	13	13	11	11
BE	10	10	10	10	10	10	12	12
CY	19	18	20	21	19	18	20	17
CZ	18	19	18	18	18	19	16	18
DK	9	9	9	9	9	9	7	6.5
EE	20	20	21	19	20	20	21	13.5

FI	14	14	14	14	14	14	10	4
FR	5	5	5	4	5	5	2	9
DE	2	4	3	3	3	3	6	6.5
EL	21	21	19	20	21	21	19	19
HU	16	16	16	17	16	16	18	16
IR	1	1	1	1	1	1	1	8
IT	15	15	15	15	15	15	15	10
NL	7	7	8	7	7	7	8	3
PL	17	17	17	16	17	17	17	21
PT	12	12	12	13	12	12	14	20
SL	3	2	2	2	2	2	4	13.5
ES	11	11	11	11	11	11	13	15
SE	6	6	6	6	6	6	5	1
CH	8	8	7	8	8	8	9	5
UK	4	3	4	5	4	4	3	2

In order to evaluate the different rankings, the values of penalty factor and the correlation coefficient are calculated. We define a penalty factor=|*-RSO ranks|, where * represents one of the seven index ranks. All penalty factors with regard to the 21 countries are calculated and their average values are shown in Table 7.

Table 7. Average penalty factor of the seven index ranking results

	E1 ranks	E2 ranks	E3 ranks	E4 ranks	E5 ranks	E6 ranks	NNRS ranks
Average	4.190	4.095	4.190	4.286	4.190	4.190	3.619

Obviously, the hybrid system integrating neural networks and rough sets is characterized by the lowest average penalty factor while the different experiments based on the NNs (E1 to E6) obtain results within the same order of magnitude. In other words, the hybrid system proposed in this paper performs best since it has the highest fit with the RSO ranking. Similar conclusions are deduced by conducting a correlation analysis of which the results are shown in Table 8.

Table 8. Correlation coefficients between the index ranking and the RSO ranking

	E1 ranks	E2 ranks	E3 ranks	E4 ranks	E5 ranks	E6 ranks	NNRS ranks
Correlation	0.652	0.650	0.631	0.641	0.643	0.645	0.716

7 Conclusions

In this paper, we presented a hybrid system which integrates neural networks and rough sets theory to combine individual road safety performance indicators into an overall composite index. In this hybrid system, two neural networks—a single neuron and a two-layer feed-forward network called a gating network—are introduced to assign weights to the seven road safety performance indicators using the Levenberg–Marquardt algorithm. However, the fact that different initial weights of the gating network and a different number of iterations possibly lead to other results thereby resulting in interpretation issues, limits the use of this technique to some extent. Therefore, rough sets theory as a useful tool in dealing with vagueness and uncertainty in decision situations is utilized to determine the most important attribute and thereby generating a decision rule to guide the network training. In our case, using the concepts of reduct and core we transformed the data into knowledge that the speed indicator should contribute most. Therefore, the initial weights of the gating network are

stressing this indicator, and the self-tuning of the neural network stops before this is no longer the case.

By simulation, optimal weights are assigned, which provide new insights into road safety policy making. More specifically, important risk domains are identified by means of relatively high weights such as speed (0.238), protective systems (0.237) and vehicle (0.204). Furthermore, final index scores are computed to assess the relative overall road safety performance of countries. Ranking results of the 21 European countries are deduced and compared with the ones calculated by the method of neural networks only. The evaluation results based on the average penalty factor and the correlation coefficients demonstrate the feasibility of this hybrid framework.

In the future, a further exploitation of this research is worthwhile. For example, a feed-forward multilayer neural network (with one or more hidden layers) could be introduced as the gating network; other features such as the degree of dependency and significance of attributes in rough sets theory are worthwhile applying. Furthermore, as the results obtained in this paper are totally derived from the data, it is meaningful to take expert knowledge into account on the one hand, and to apply the methodology on another road safety performance indicator data set on the other hand.

References

Atiya A, Ji CY (1997) How initial conditions affect generalization performance in large networks. IEEE Trans Neural Netw 8(2):448–451

Cao G, Shiu SCK, Wang X (2003) A fuzzy-rough approach for the maintenance of distributed case-based reasoning systems. Soft Comput 7(8):491–499

European Transport Safety Council (2001) Transport Safety Performance Indicators, ETSC, Brussels

European Transport Safety Council (2008) Countdown to 2010, only two more years to act, 2nd Road Safety PIN Report ETSC, Brussels

European Commission (2001) White Paper European transport policy for 2010: time to decide, Commission of the European Communities, Brussels

Hagan MT, Menhaj MB (1994) Training feed forward network with the Marquardt algorithm. IEEE Trans Neural Netw 5(6):989–993

Haykin S (1999) Neural networks: a comprehensive foundation. Macmillan College Publishing Company, New York

Hermans E (2009) A methodology for developing a composite road safety performance index for cross-country comparison. PhD Thesis, Hasselt university, Belgium

Hermans E, Ruan D, Brijs T, Wets G, Vanhoof K (2008a) Evaluation of road safety performance indicators using OWA operators. Comput Eng Inform Sci 1:695–700

Hermans E, Van den Bossche F, Wets G (2008b) Combining road safety information in a performance index. Accid Anal Prev 40:1337–1344

Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. Neural Netw 2:359–366

Jiang Y, Zhang H, Xie J, Meng K (2006) An estimation model of research cost based on rough set and artificial neural network. LNAI 4114:766–771

Li T, Ruan D, Wets G, Song J, Xu Y (2007) A rough sets based characteristic relation approach for dynamic attribute generalization in data mining. Knowl-Based Syst 20(5):485–494

Litman T (2007) Developing indicators for comprehensive and sustainable transport planning. Transportation Research Record 2017, Transportation Research Board, 10–15

Luo J, Shao H (2006) Developing soft sensors using hybrid soft computing methodology: a neurofuzzy system based on rough set theory and genetic algorithms. Soft Comput 10(1):54–60

Midelfart H, Komorowski HJ, Nørsett K, Yadetie F, Sandvik AK, Lægreid A (2002) Learning rough set classifiers from gene expressions and clinical data. Fundamenta Informaticae 53:155–183

Nardo M, Saisana M, Saltelli A, Tarantola S, Hoffman A, Giovannini E (2005) Handbook on constructing composite indicators: methodology and user guide. OECD Statistics Working Papers, Organisation for Economic Cooperation and Development, Paris

Øhrn A, Komorowski J, Skowron A, Synak P (1998) The design and implementation of a knowledge discovery toolkit based on rough sets: the ROSETTA system. In: Polkowski L, Skowron A (eds) Rough sets in knowledge discovery. 1. Methodoloy and applications. Studies in fuzziness and soft computing, vol 18. Physica-Verlag, Heidelberg, pp 376–399

Pawlak Z (1991) Rough sets. Theoretical aspects of reasoning about data. Kluwer Academic Publishers, Boston

Pawlak Z, Polkowski L, Skowron A (2005) Rough sets. In: Rivero LC, Doorn JH, Ferragine VE (eds) Encyclopedia of database technologies and applications, Idea Group Reference, pp 575–580

Pawlak Z, Skowron A (2007) Rudiments of rough sets. Inform Sci 177(1):3–40

Peters JF, Skowron A (2002) A rough set approach to knowledge discovery. Int J Intell Syst 17(2):109–112

Reich Y, Baraib SV (2000) A methodology for building neural networks models from empirical engineering data. Eng Appl Artif Intell 13:685–694

Rumelhart DE, Hinton GE, Wiliams RJ (1986) Learning representations by back-propagating errors. Nature 323:533–536

SafetyNet Work Package 3 (2005) State-of-the-art Report on Road Safety Performance Indicators, SafetyNet

SARTRE 3 Report (2004) European Drivers and Road Risk, SARTRE

Shen Y, Hermans E, Ruan D, Wets G, Vanhoof K, Brijs T (2008) Development of a composite road safety performance index based on neural networks. In: Proceedings of 2008 international conference on intelligent systems and knowledge engineering, Xiamen, IEEE Press 2:901–906

Skowron A (1995) Extracting laws from decision tables: a rough set approach. Comput Intell 11:371–388

Swiniarski RW (2001) Rough set methods in feature reduction and classification. Int J Appl Math Comput Sci 11(3):565–582

Swiniarski RW, Skowron A (2003) Rough set methods in feature selection and recognition. Pattern Recognit Lett 24(6):833–849

Wegman F, Commandeur J, Doveh E, Eksler V, Gitelman V, Hakkert S, Lynam D, Oppe S (2008) SUNflowerNext: towards a composite road safety performance index, Deliverable D6.16 of the EU FP6 project SafetyNet

Werbos PJ (1988) Back-propagation: past and future. In: Proceeding of international conference on neural networks, San Diego, CA, 1:343–354

Witten IH, Frank E (2005) Data mining: practical machine learning tools and techniques, 2nd edn. Morgan Kaufmann, San Francisco

Xiang C, Ding SQ, Lee TH (2005) Geometrical interpretation and architecture selection of MLP. IEEE Trans Neural Netw 16(1):84–96

Xiao Z, Ye S, Zhong B, Sun C (2009) BP neural network with rough set for short term load forecasting. Expert Syst Appl 36:273–279