Expert Systems with Applications xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Permanent disability classification by combining evolutionary Generalized Radial Basis Function and logistic regression methods

A. Castaño [a], Francisco Fernández-Navarro [b,*], P.A. Gutiérrez [b], César Hervás-Martínez [b]

[a] Department of Informatics, University of Pinar del Río, Cuba
[b] Department of Computer Science and Numerical Analysis, University of Córdoba, Campus de Rabanales, Albert Einstein Building, 3rd Floor, 14071 Córdoba, Spain

ARTICLE INFO

Keywords:
Neural networks
Multi-classification
Logistic regression
Evolutionary algorithms
Permanent disability classification

ABSTRACT

Recently, a novelty multinomial logistic regression method where the initial covariate space is increased by adding the nonlinear transformations of the input variables given by Gaussian Radial Basis Functions (RBFs) obtained by an evolutionary algorithm was proposed. However, there still exist some problems with the standard Gaussian RBF, for example, the approximation of constant valued functions or the approximation of high dimensionality associated to some real problems. In order to face these problems, we propose the use of the generalized Gaussian RBF (GRBF) instead of the standard Gaussian RBF. Our approach has been validated with a real problem of disability classification, to evaluate its effectiveness. Experimental results show that this approach is able to achieve good generalization performance.

1. Introduction

In artificial neural networks (ANNs), the hidden neurons are the functional units and can be considered as generators of function spaces. Most existing neuron models are based on the summing operation of the inputs, and, more particularly, on sigmoidal unit functions, resulting in what is known as the Multilayer Perceptron (MLP). However, alternatives to MLP emerged in the last few years: Product Unit Neural Network (PUNN) models are an alternative to MLPs and are based on multiplicative neurons instead of additive ones. They correspond to a special class of feed-forward neural network introduced by Durbin and Rumelhart (1989). While MLP network models have been very successful, networks that make use of Product Units (PUs) have the added advantage of increased information capacity (Durbin & Rumelhart, 1989). That is, smaller PUNNs architectures can be used rather than those used with MLPs (Ismail & Engelbrecht, 2002). They aim to overcome the non-linear effects of variables by means of non-linear basis functions, constructed with the product of the inputs raised to arbitrary powers. These basis functions express possible strong interactions between the variables, where the exponents may even take on real values and are suitable for automatic adjustment.

Another interesting alternative to MLPs are Radial Basis Function Neural Networks (RBFNNs). RBFNNs can be considered a local approximation procedure, and the improvement in both its approximation ability, as well as in the construction of its architecture has been noteworthy (Bishop, 1991). RBFNNs have been used in the most varied domains, from function approximation to pattern classification, time series prediction, data mining, signals processing, health monitoring, and non-linear system modelling and control (Howlett & Jain, 2001; Zheng, Li, & Wang, 2011). RBFNNs use, in general, hyper-ellipsoids to split the pattern space. In many cases, MLP, PU and RBF networks are trained by using evolutionary algorithms (EAs), thus obtaining advantages with respect to traditional training approaches (Chakravarty & Dash, 2011; Fernández-Navarro, Hervás-Martínez, Cruz, Gutierrez, & Valero, 2011a; Fernández-Navarro, Hervás-Martınez, Gutierrez, & Carboreno, 2011d; Tallón-Ballesteros & Hervás-Martínez, 2011; Yao, 1999).

On the other hand, logistic regression (LR) has become a widely used and accepted method of analysis of binary or multi-class outcome variables as it is more flexible and it can predict the probability of the state of a multi-class variable based on the predictor variables. Guti'errez, Hervás-Martínez, and Martínez-Estudillo (2011) proposed a multinomial logistic regression method, combining evolutionary Radial Basis Function (ERBF) and LR methods. The LR methods apply a logit function to the linear combination of the input variables. The coefficients values of each input variable are estimated by means of the Iterative Reweighted Least Square (IRLS) algorithm. Roughly, the methodology is divided into 3 steps. Firstly, an evolutionary algorithm (EA) is applied to estimate the parameters of the RBF. Secondly, the input space is increased by adding the nonlinear transformation of the input variables given by the RBFs of the best individual in the last generation of the EA. Finally, the LR algorithms are applied in this new covariate space.

The standard Gaussian RBF has some drawbacks, for example, its performance decreases drastically when it is applied to approximate constant valued function or when dimensionality grows. For

* Corresponding author. Tel.: +34 957 21 83 49; fax: +34 957 21 83 60.
E-mail address: i22fenaf@uco.es (F. Fernández-Navarro).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.01.186

this reason, we propose the use of a Generalized RBF (GRBF) (Castaño, Fernández-Navarro, Hervás-Martínez, Gutierrez, & García, 2010; Fernández-Navarro, Hervás-Martínez, Sánchez-Monedero, & Gutierrez, in press), instead of the standard Gaussian RBF. This novelty basis function incorporates a new parameter, τ, that allows the contraction–relaxation of the standard RBF, solving the problems previously stated.

The performance of the proposed multinomial logistic regression methodology was evaluated in a real problem of permanent disability classification. Permanent disability is a term used in the insurance industry and law. Generally speaking, it means that due to a sickness or injury a person is unable to work in their own, or any occupation for which they are suited by training, education, or experience. In Spain, the evaluation and classification of permanent disability follows a procedure which is clearly defined and divided into three development phases: introduction, instruction and resolution.

The main principles of the measures adopted with the aim of obtaining a consolidated and rationalized system for the determination of permanent disability are the contributory element, equity and solidarity. Furthermore, in order to establish greater legal security in the process of determining permanent disability, it is necessary to elaborate a list of diseases and the evaluation of their influence on the reduction of work capacity. This list must be created according to objective criteria based on the actual evaluations and proceedings of the *disability assessment teams*.

To understand the nature of *permanent disability*, it is necessary to define the terminology first. Permanent disability takes into account continuous alteration of health and its impact on the worker's occupational situation. The *disability assessment team* is supported by a *medical unit*. The medical unit's competencies are: to examine the disability situation of the worker, to determine the reduction or alteration of the physical integrity of the worker, to determine the level of incapacity for work, to determine whether the character of the disease is common or professional, to extend the period of medical observation in case of professional diseases, to monitor programs for the control of temporal disability compensations, and to provide technical assistance and advice on any contentious issues concerning occupational disabilities.

In our work we consider three main categories that can be assigned to a worker depending on the degree of permanent disability: *no disability* (when the worker is not assigned the status of permanent disability), *permanent disability* (when the worker is assigned some degree of permanent disability) and *fee* (when the worker is not assigned any degree of permanent disability, but is financially compensated). The objective of this study is to offer an initial model based on artificial neural networks and logistic regression which facilitates preparing reports in the process of determining the existence of permanent disability. This model allows to obtain an approximation of the expected result for each case of permanent disability. The training dataset used to obtain the model is composed of information from reports of the medical unit. Each report is tagged with one of the three categories (no disability, permanent disability or fee). An important characteristic of the dataset is that it is highly unbalanced.

2. Occupational situation and permanent disability

Permanent disability (PD), in its contributory modality, takes into account the continuous alteration of health and, particularly, its impact on occupational situation.

It has an exclusively professional profile and its evaluation should avoid references to other circumstances, such as socio-economic status, age, family, etc. These circumstances may be considered in order to evaluate other effects, but should not be taken into account when determining the degree of disability to be protected by contributory income.

The occupational situations to be protected by the status of permanent disability are:

- Permanent disability which, in practice, stands for the lack of income due to the loss of salary which is a result of either temporary, or permanent disability. This lack of income is alleviated by financial aid.
- The necessity to recover psycho-physical well being.
- The necessity to receive financial support during the process of recovery.
- The process of reintegrating a disabled person into work environment, which should be protected by selective employment.

Depending on the determining cause, permanent disability is classified according to the following degrees:

- *Partial PD* for usual occupation means that a worker's capacity to perform his/her job is diminished by not less than 33%. However, it does not prevent him/her from performing tasks which are fundamental for his/her occupation.
- *Total PD* for usual occupation means that a worker is unable to perform tasks which are fundamental for his/her occupation, but may opt for a different occupation.
- *Absolute PD* means that a worker is unable to perform any profession.
- *Grand disability* means that a worker who is affected by PD due to his/her physical and functional impairments requires assistance in basic life activities such as dressing up, moving from one place to another, eating, etc.
- Non-disabling permanent damages refers to permanent impairments which do not have impact on work capacity, but mean that a worker's physical integrity is reduced. Non-disabling permanent damages are classified by "Ley General de la Seguridad Social".

In case of accidents, whether work accidents or not, the term "usual occupation" should be understood as work performed by a worker at the time of the accident.

2.1. Initial data and variables

The medical unit of the disability assessment team elaborates synthesis medical reports (SMR) to evaluate permanent disability. We use these reports as a source of information for our experiments. Synthesis medical reports are based on:

1. Clinical examination performed by a medical evaluator.
2. Medical reports provided by the ill.
3. Complementary tests and examinations requested by the medical evaluator.

The data used here had been obtained from the synthesis medical reports and proceedings of the sessions held by the disability assessment team which were then compiled into files. Some data, like age or sex, have been extracted directly from these documents while others, like occupational repercussion, have been collected by qualified persons.

For each file there have been obtained the following attributes:

- From the synthesis medical reports: Age, sex, occupation, sick leave period, diseases.
- From the proceedings of the sessions held by the disability assessment team: Classification (permanent disability degree), contingency, period of time between examinations.
- Occupational repercussion. The following information has been taken into account when evaluating it as low, middle or high:

Q1 Please cite this article in press as: Castaño, A., et al. Permanent disability classification by combining evolutionary Generalized Radial Basis Function and logistic regression methods. *Expert Systems with Applications* (2012), doi:10.1016/j.eswa.2012.01.186

– Functional repercussion of different diseases.
– Worker's occupation.

The classification (permanent disability degree) is grouped into:

- No disability (ND).
- Permanent disability (PD).
- Fee (F).

The contingency can be classified into two types:

- Common
  - Common disease (CD).
  - Non-working accident (NWA).
- Professional
  - Occupational disease (OD).
  - Working accident (WA).

We have used the code of the Spanish ''National Classification of Occupations'' (CNO-94) to collect the data related to professions. To gather the data related to diseases, we have used the ''International Classification of Diseases'' (ICD9-CM).

The final variables used in our work are shown in Table 1.

A total of 978 records have been extracted from the data between 2002 and 2003.

3. Generalized Radial Basis Function

A RBF is a function which has been built taking into account a distance criterion with respect to a center. Different basis functions like multiquadratic functions, inverse multiquadratic functions and Gaussian functions have been proposed, but normally the selected one is the Gaussian function. The standard RBF model is described as follows:

$$B_j(\mathbf{x},\mathbf{w}_j) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{c}_j\|}{r_j}\right)^2. \quad (1)$$

where $\mathbf{w}_j = (\mathbf{c}_j, r_j)$, $\mathbf{c}_j = (c_{j1}, c_{j2}, \ldots, c_{jk})$ is the center or average of the jth Gaussian RBF transformation, r_j is the corresponding radius or standard deviation.

In the same way that the Gaussian RBF is based on the Gaussian distribution, we could obtain different RBFs considering parametric versions of the Gaussian distribution. One example of a parametric version of the Gaussian distribution is the Generalized Gaussian distribution (Andai, 2009; Nandi & Mämpel, 1995; Sharifi & Leron-Garcia, 1995). This distribution function adds a real parameter, τ, allowing the representation of different distribution functions, like the Laplacian distribution for $\tau = 1$ or the uniform distribution for $\tau \to 0$.

Based on this distribution, we define the Generalized RBF by replacing the quadratic exponent of previous model by τ:

$$B_j(\mathbf{x},\mathbf{w}_j) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{c}_j\|}{r_j}\right)^{\tau}, \quad (2)$$

In this case $\mathbf{x}$ also includes the parameter τ_j representing the exponent of the basis function, where $c_{ji}, \tau_j, r_j \in \mathbb{R}$. Fig. 1 presents the radial unit activation for the GRBF for different values of τ.

4. Neuro-logistic models

In the classification problem, some measurements x_i, $i = 1, 2, \ldots, k$ are taken on a single pattern, and the patterns are classified into one of J populations. The measurements x_i are random observations from these J classes. A training sample $D = \{(\mathbf{x}_n, \mathbf{y}_n);\ n = 1, 2, \ldots, N\}$ is available, where $\mathbf{x}_n = (x_{1n}, \ldots, x_{kn})$ is the vector of measurements taking values in $\mathbf{\Omega} \subset \mathbb{R}^k$, and $\mathbf{y}_n$ is the class level of the nth individual.

Table 1
List of variables and associated description of the dataset obtained from the synthesis medical reports and proceedings of the sessions held by the disability assessment team.

Variable	Description
x_1	Age
x_2	Sex
x_{3-21}	CNO-94
x_{22}	Sick leave time
x_{23-42}	Principal categories of ICD9-CM
x_{43}	Low occupational repercussion
x_{44}	Middle occupational repercussion
x_{45}	High occupational repercussion
x_{46}	Total number of diseases
x_{47}	CD contingency
x_{48}	NWA contingency
x_{49}	OD contingency
x_{50}	WA contingency
x_{51}	Period of time between examinations
Class	Description
ND	No disability
PD	Permanent disability
F	Fee

The common technique of representing the class levels using a ''1-of-J'' encoding vector is adopted, $\mathbf{y} = (y^{(1)}, y^{(2)}, \ldots, y^{(J)})$, such as $y^{(l)} = 1$ if $\mathbf{x}$ corresponds to an example belonging to class l and $y^{(l)} = 0$ otherwise.

Based on the training sample, we wish to find a decision function $F : \mathbf{\Omega} \to \{1, 2, \ldots, J\}$ for classifying the individuals. In other words, F provides a partition, say $D_1, D_2, \ldots, D_J$, of $\mathbf{\Omega}$, where D_l corresponds to the lth class, $l = 1, 2, \ldots, J$, and measurements belonging to D_l will be classified as coming from the lth class. A misclassification occurs when the decision rule F assigns an individual (based on the measurement vector) to a class j when it is actually coming from a class $l \neq j$.

Logistic Model supposes that the conditional probability that $\mathbf{x}$ belongs to class l verifies: $p(y^{(l)} = 1|\mathbf{x}) > 0$, $l = 1, 2, \ldots, J$, $\mathbf{x} \in \mathbf{\Omega}$, and sets the function:

$$f_l(\mathbf{x},\boldsymbol{\theta}_l) = \log\frac{p(y^{(l)} = 1|\mathbf{x})}{p(y^{(J)} = 1|\mathbf{x})}, \quad (3)$$

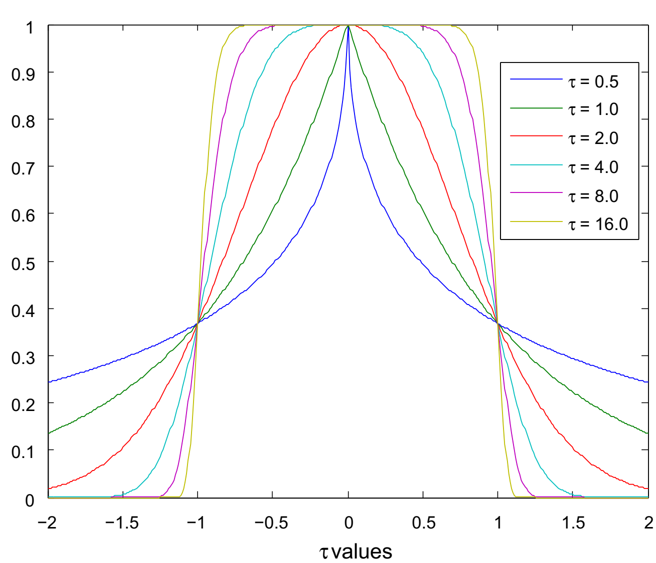

Fig. 1. Radial unit activation in one-dimensional space with $c = 0$ and $r = 1$ for the GRBF with different values of τ.

where θ_l is the weight vector corresponding to class l, and $f_J(\mathbf{x}, \theta_J) = 0$. Under a multinomial logistic regression, the probability that $\mathbf{x}$ belongs to class l is then given by:

$$p(y^{(l)} = 1|\mathbf{x}, \boldsymbol{\theta}) = \frac{\exp f_l(\mathbf{x}, \boldsymbol{\theta}_l)}{\sum_{j=1}^{J} \exp f_j(\mathbf{x}, \boldsymbol{\theta}_j)}, \quad l = 1, 2, \ldots, J, \tag{4}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{J-1})$. The hybrid neuro-logistic models are based on the combination of the standard linear model and nonlinear terms constructed with RBFs or GRBFs, which captures possible locations in the covariate space. The general expression of the model is given by:

$$f_l(\mathbf{x}, \boldsymbol{\theta}_l) = \alpha_0^l + \sum_{i=1}^{k} \alpha_i^l x_i + \sum_{j=1}^{m} \beta_j^l B_j(\mathbf{x}, \mathbf{w}_j) \tag{5}$$

where $l = 1, 2, \ldots, J-1$, $\boldsymbol{\theta}_l = (\boldsymbol{\alpha}^l, \boldsymbol{\beta}^l, \mathbf{W})$ is the vector of parameters for each discriminant function, $\boldsymbol{\alpha}^l = (\alpha_0^l, \alpha_1^l, \ldots, \alpha_k^l)$ and $\boldsymbol{\beta}^l = \left(\beta_1^l, \ldots, \beta_m^l\right)$ are the coefficients of the multilogistic regression model and $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m)$ are the parameters of the nonlinear transformations and B_j is the RBF or GRBF (described in Section 3).

The general structure of this kind of models can be analyzed in Fig. 2.

5. Estimation of neuro-logistic parameters

In the supervised learning context, the components of the weight vectors $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{J-1})$ are estimated from the training dataset D. To perform the maximum likelihood estimation of $\boldsymbol{\theta}$, one can minimize the negative log-likelihood function:

$$\begin{aligned} L(\boldsymbol{\theta}) &= -\frac{1}{N} \sum_{n=1}^{N} \sum_{l=1}^{J} \left(y_i^l \log p(\mathbf{y}_n|\mathbf{x}_n, \boldsymbol{\theta})\right) \\ &= \frac{1}{N} \sum_{n=1}^{N} \left[-\sum_{l=1}^{J} y_n^{(l)} f_l + \log \sum_{l=1}^{J} \exp f_l \right], \end{aligned} \tag{6}$$

where $f_l = f_l(\mathbf{x}_n, \boldsymbol{\theta}_l)$ corresponds to the hybrid model defined in (5).

The methodology proposed tries to maximize the log-likelihood function where classical gradient methods are not recommended due to the convolved nature of the error function. It is based on the combination of an evolutionary programming algorithm (EP) (global explorer) and a local optimization procedure (local exploiter) carried out by the standard maximum likelihood optimization method.

In this paper, two different algorithms have been considered for obtaining the maximum likelihood solution for the multilogistic regression model, both available in the WEKA workbench (Witten & Frank, 2005): MultiLogistic and SimpleLogistic. The first one is an algorithm for building a multinomial logistic regression with a ridge estimator to prevent overfitting by penalizing large coefficients. This model is trained with a Quasi-Newtonian Method. The second one builds a multinomial logistic regression model fitting the coefficients with the LogitBoost algorithm (Landwehr, Hall, & Frank, 2005).

The estimation of the model coefficients is divided into three steps.

Step 1. We apply an EP algorithm to find the basis functions:

$$\mathbf{B}(\mathbf{x}, \mathbf{W}) = \{B_1(\mathbf{x}, \mathbf{w}_1), B_2(\mathbf{x}, \mathbf{w}_2), \ldots, B_m(\mathbf{x}, \mathbf{w}_m)\}, \tag{7}$$

corresponding to the nonlinear part of $f(\mathbf{x}, \boldsymbol{\theta}_l)$. We have to determine the number of basis functions m and the weight matrix $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m)$.
The weight matrix $\mathbf{W}$, the parameters of the output layer ($\boldsymbol{\beta}$ vector) and the structure of the GRBF are estimated by means of an evolutionary neural network algorithm that optimizes the error function given by the negative log-likelihood for N observations associated with the neural network model (see Eq. (6)). The specific details of this EP algorithm can be found in some previous works (Fernández-Navarro, Hervás-Martínez, García-Alonso, & Torres-Jimenez, 2011b, 2011c, 2012).
As we discussed previously, the model introduces a new parameter, τ, which it is necessary to be estimated during the evolutionary process. In the initialization step of the EP, the τ value of all basis function is set to 2, since the GRBF with $\tau = 2$ is equivalent to the standard Gaussian RBF. On the other hand, the parametric mutator modified the τ parameter of each basis function by adding an uniform random value ζ in the interval $[-0.25, 0.25]$. Finally, when the structural mutator adds a new GRBF hidden node, it is included in the model with a $\tau = 2$.
We only consider the estimated weight matrix $\widehat{\mathbf{W}} = (\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2, \ldots, \hat{\mathbf{w}}_m)$, which builds the basis functions. The values for the $\boldsymbol{\beta}$ vector will be determined in step 3 together with those of the $\boldsymbol{\alpha}$ coefficient vector.

Step 2. We consider the following transformation of the input space by including the nonlinear basis functions obtained by the EP algorithm in step 1:

$$\begin{aligned} &H : \mathbb{R}^k \to \mathbb{R}^{k+m}, \\ &(x_1, x_2, \ldots, x_k) \to (x_1, x_2, \ldots, x_k, z_1, \ldots, z_m), \end{aligned} \tag{8}$$

where $z_1 = B_1(\mathbf{x}, \hat{\mathbf{w}}_1), \ldots, z_m = B_m(\mathbf{x}, \hat{\mathbf{w}}_m)$.

Step 3. In the third step, we minimize the negative log-likelihood function for N observations:

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{N} \sum_{n=1}^{N} \left[-\sum_{l=1}^{J} y_n^{(l)} \Upsilon + \log \sum_{l=1}^{J} \exp \Upsilon \right], \tag{9}$$

where $\Upsilon = (\boldsymbol{\alpha}^l \mathbf{x}_n + \boldsymbol{\beta}^l \mathbf{z}_n)$, $\mathbf{x}_n = (1, x_{1n}, \ldots, x_{kn})$ and $\mathbf{z}_n = (z_{1n}, \ldots, z_{mn})$. Now, the Hessian matrix of the negative log-likelihood in the new variables $x_1, x_2, \ldots, x_k, z_1, \ldots, z_m$ is semi-definite positive. The estimated coefficient vector $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \widehat{\mathbf{W}})$ determines the model of (5) with $B_j(\mathbf{x}, \mathbf{w}_j)$ defined as (2).

In this final step, both logistic regression algorithms have been used for obtaining the parameter matrix $\boldsymbol{\theta}$. Moreover, two different versions of the hybrid neuro-logistic models have been considered: LR models with only the non-linear part, i.e. the model does not include the initial covariates of the problem, and LR models with both the linear and the non-linear part, i.e., the models. The combined application of both algorithms logistic regression with the two evolutionary algorithms (using RBF and GRBF) with out initial covariates results into four different methods: MultiLogistic regression with GRBFs (MLGRBF), SimpleLogistic regression with GRBFs (SLGRBF), MultiLogistic regression whith RBFs (MLRBF) and SimpleLogistic regression with RBFs (SLRBF). In the same way other four methods are obtained including initial variables: MLIGRBF, SLIGRBF, MLIRBF and SLIRBF.

6. Experiments

6.1. Experimental design and statistical analysis

Various methods discussed above were compared to the following state-of-art algorithms (since they are some of the best performing algorithms of recent literature on classification problems):

- The k Nearest Neighbour (k-NN) classifier, adjusting the value of k using a nested 10-fold cross-validation.
- A Gaussian Radial Basis Function Network (RBFNetwork) available in the WEKA workbench (Witten & Frank, 2005).
- Both standard logistic regression algorithms presented in Section 5: SimpleLogistic (SLogistic) and MultiLogistic (MLogistic).

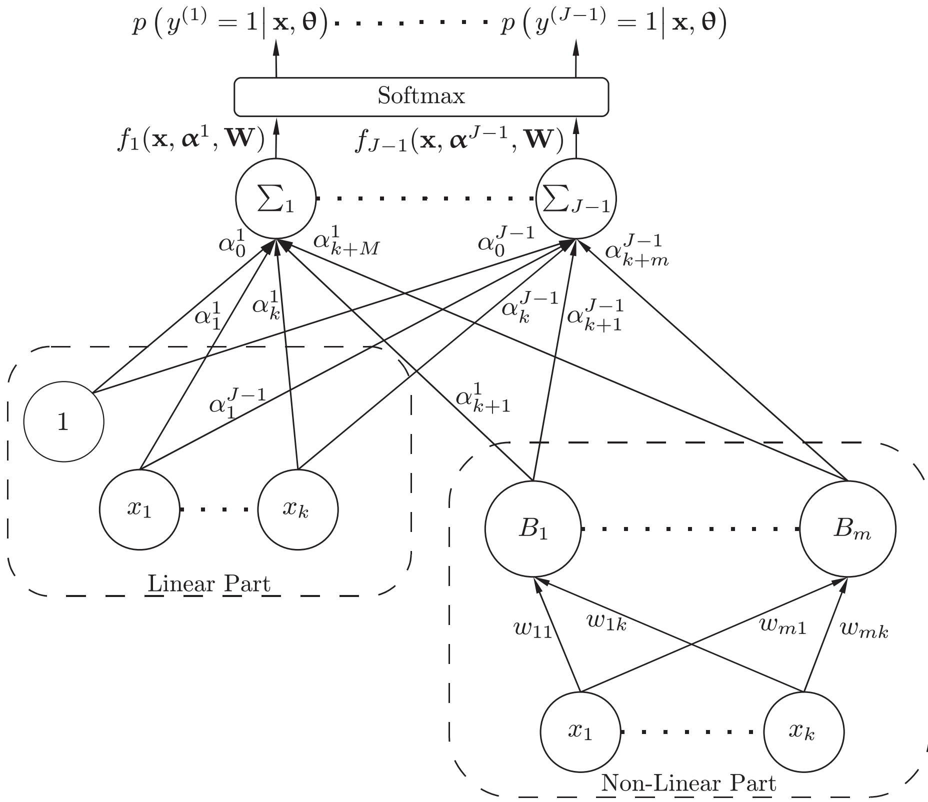

Fig. 2. Structure of neuro logistic models.

- The Naive Bayes standard learning algorithm (NaiveBayes) (Witten & Frank, 2005).

A 10-fold cross-validation has been applied and the performance has been evaluated by using the Correct Classification Rate or accuracy (C) in the generalization set (C_G). When applying the algorithms proposed (GRBF and RBF (Guti'errez et al., 2011) methods), ten repetitions are performed per each fold, and when applying the rest of methods, the 10-fold process is repeated ten times, in order to obtain an average and a standard deviation of the C_G from the same sample size (100 models). A simple linear rescaling of the input variables was performed in the interval [−2,2], X_i^* being the transformed variables, for RBFs (Guti'errez et al., 2011) and GRBF methodologies.

Table 2 shows in the second column the results obtained with the different techniques tested. The SLIGRBF method obtained the best result in terms of C_G out of all the techniques compared. Other important observation is that GRBF methods generally outperform their RBF equivalents, obtaining also a lower standard deviation. It is well known that Neural Networks, Evolutionary Computations, and Fuzzy Logics, are three representative methods of Soft Computing (Corchado, Arroyo, & Tricio, in press). In this paper, we hybridize two of them (Neural Networks and Evolutionary Computation). Therefore, we could consider our proposal as a competitive method within the scope of Soft Computing.

In order to ascertain the statistical significance of the observed differences between the mean C_G of the best models obtained for each methodology, we have applied the Mann–Whitney U rank sum test for all pairs of algorithms since a previous evaluation of the Kolmogorov–Smirnov test (KS-test) stated that a normal distribution cannot be assumed in all the results reported by the algorithms and the non-parametric Kruskal–Wallis test concluded that these differences were significant. The results of the Mann–Whitney U rank sum test are included in Table 2 column 3–5. From the analysis of these results, the SLIGRBF method has to be highlighted as the most competitive one (with only one draw), followed by SLIRBF. Consequently, GRBFs are better suited for classifying permanent disability than RBFs.

One of the major advantages of the SLIGRBF model is the reduced number of features and GRBFs included in the final expression, since the MA reduces its complexity by pruning mutations and the Simple Logistic algorithm does feature selection reliably. This can result in a better interpretability of the model, which is especially important when dealing with real problems. In this way, Table 3 includes the best predictor functions of the SLIGRBF

Table 2
Mean, standard deviation, maximum and minimum values of the accuracy results (C_G) from 100 executions of a 10-fold cross validation. Number of wins, draws and loses when comparing the different methods using the Mann–Whitney U rank sum test $\alpha = 0.05$. Q3

	C_G (%)	Mann–Whitney U test		
	Mean ± SD	# Wins	# Draws	# Loses
EGRBF	85.26 ± 5.08	5	4	5
MLGRBF	85.76 ± 5.42	5	5	4
SLGRBF	85.30 ± 4.90	5	5	4
MLIGRBF	89.03 ± 3.34	11	1	2
SLIGRBF	**90.70 ± 3.02**	**13**	**1**	**0**
ERBF	79.76 ± 11.36	1	2	11
MLRBF	79.88 ± 11.20	1	2	11
SLRBF	79.56 ± 13.54	1	2	11
MLIRBF	86.39 ± 8.96	5	5	4
SLIRBF	89.86 ± 9.40	12	2	0
k-NN	66.04 ± 8.12	0	0	14
RBFNetwork	86.75 ± 9.30	6	4	4
SLogistic	89.77 ± 9.39	11	2	1
MLogistic	86.54 ± 9.31	5	5	4
NaiveBayes	84.17 ± 9.15	4	0	10

Table 3
Probability expression of the best SLIGRBF model, C_G and test confusion matrix.

Best *SLIGRBF* permanent disability probability model

$$p_{ND} = \frac{e^{f_{PD}(\mathbf{x},\theta)}}{1+e^{f_{PD}(\mathbf{x},\theta)}+e^{f_{ND}(\mathbf{x},\theta)}};\ p_{PD} = \frac{e^{f_{ND}(\mathbf{x},\theta)}}{1+e^{f_{PD}(\mathbf{x},\theta)}+e^{f_{ND}(\mathbf{x},\theta)}}$$

$$f_{ND}(\mathbf{x},\theta) = 0.78 + 0.23x_1 + 0.25x_{22} - 0.19x_{32} + 0.27x_{35} + 1.05x_{42} - 1.16x_{47} + 0.67x_{49} + 6.76GRBF_1$$

$$f_{PD}(\mathbf{x},\theta) = 5.16 + 0.30x_{10} + 0.25x_{22} + 0.19x_{32} + 0.27x_{35} + 1.05x_{42} - 1.38x_{47} - 0.72x_{49} - 3.76GRBF_1$$

$$GRBF_1 = \exp\left(-\frac{((x_2+0.07)^2+(x_{35}+0.10)^2+(x_{41}+1.98)^2+(x_{42}-0.73)^2+(x_{43}-1.41)^2)^{0.5}}{5.20}\right)^{17.28}$$

$x_1 \leftarrow$ (age); $x_{10} \leftarrow$ (rcno94 = 12); $x_{22} \leftarrow$ (sick leave time)
$x_{32} \leftarrow$ (disease10); $x_{35} \leftarrow$ (disease13); $x_{41} \leftarrow$ (disease19)
$x_{42} \leftarrow$ (disease20); $x_{43} \leftarrow$ (low occupational repercussion); $x_{47} \leftarrow$ (contingency = CD); $x_{49} \leftarrow$ (contingency = WA)

$x_i \in [-2.0, 2]$; $C_G = 96.43\%$

Generalization confusion matrix

	Predicted		
Target	*NI*	*I*	*B*
NI	46	2	0
I	1	32	0
B	0	0	3

model obtained for the Permanent Disability classification problem. The model is formed only for ten input variables, demonstrating the reliability of both the evolutionary algorithm and the Simple Logistic algorithm to effectively reduce the feature space.

7. Conclusions

We have study the combination of Evolutionary Generalized Radial Basis Function instead of Evolutionary Radial Basis Function and logistic regression methods. This basis function solve some problems that lacks the performance of the standard Gaussian model, such as the approximation of constant valued function or the approximation of high dimensionality datasets. The good synergy between these two techniques has been experimentally proved using a permanent disability classification problem.

The hybrid neuro-logistic models have proved to serve as an accurate tool in the classification of permanent disability. A comparative study between an extensive collection of standard classifiers and the results of the statistical tests applied, and the hybrid neuro-logistic models shows that the latter are more precise in determining the degree of permanent disability.

Our hybrid models include a non-linear component (from different kinds of neural networks) and a standard linear component, combining both in a logistic regression predictor. The complexity of the model and the high amount of parameters involved in these classifiers encouraged us to use a combined methodology, including an evolutionary algorithm and a standard maximum-likelihood optimization process.

Useful information could be extracted from the most accurate model, given its simple structure (number of connections and number of hidden neurons). Simple structure is one of the main advantages of the models presented.

The obtained model is not intended to be a widely used tool in the classification of permanent disability. First, it would be necessary to examine more data as the scope of the PD problem is very broad due to the high number and complexity of cases. However, our findings can be used to develop new, improved systems. For instance, an extended model could be used to create an information system, both for patients and professionals, which would provide assistance in the evaluation of permanent disability.

Acknowledgement

This work has been partially subsidized by the TIN 2008-06681-C06-03 project of the Spanish Inter-Ministerial Commission of Science and Technology (MICYT), FEDER funds and the P08-TIC-3745 project of the "Junta de Andalucía" (Spain). The research of Francisco Fernández-Navarro has been funded by the "Junta de Andalucía" Predoctoral Program, grant reference 390015-P08-TIC-3745. This work has been partially subsidized with the project "Doctoral Training on Softcomputing" supported by the Junta de Andalucía, the Ibero-American University Postgraduate Association (AUIP) and the Ministry of Higher Education of the Republic of Cuba.

References

Andai, A. (2009). On the geometry of generalized gaussian distributions. *Journal of Multivariate Analysis, 100*(4), 777–793.
Bishop, C. M. (1991). Improving the generalization properties of radial basis function neural networks. *Neural Computation, 3*(4), 579–581.
Castaño, A., Fernández-Navarro, F., Hervás-Martínez, C., Gutierrez, P. A., & García, M. M. (2010). Classification by evolutionary generalized radial basis functions. *International Journal of Hybrid Intelligent Systems, 7*(1), 1–10.
Chakravarty, S., & Dash, P. K. (2011). Dynamic filter weights neural network model integrated with differential evolution for day-ahead price forecasting in energy market. *Expert Systems with Applications, 38*(9), 10974–10982.
Corchado, E., Arroyo, A., & Tricio, V. (in press). Soft computing models to identify typical meteorological days. *Logic Journal of IGPL*. Q2
Durbin, R., & Rumelhart, D. (1989). Products units: A computationally powerful and biologically plausible extension to backpropagation networks. *Neural Computation, 1*(1), 133–142.
Fernández-Navarro, F., Hervás-Martínez, C., Cruz, M., Gutierrez, P. A., & Valero, A. (2011a). Evolutionary q-gaussian radial basis function neural network to determine the microbial growth/no growth interface of *Staphylococcus aureus*. *Applied Soft Computing, 11*(3), 3012–3020.
Fernández-Navarro, F., Hervás-Martínez, C., García-Alonso, C., & Torres-Jimenez, M. (2011b). Determination of relative agrarian technical efficiency by a dynamic over-sampling procedure guided by minimum sensitivity. *Expert Systems with Applications, 38*(10), 12483–12490.
Fernández-Navarro, F., Hervás-Martínez, C., & Gutíerrez, P. (2011c). A dynamic over-sampling procedure based on sensitivity for multi-class problems. *Pattern Recognition, 44*(8), 1821–1833.
Fernández-Navarro, F., Hervás-Martínez, C., Gutierrez, P. A., & Carboreno, M. (2011d). Evolutionary q-gaussian radial basis functions neural networks for multi-classification. *Neural Networks, 24*(7), 779–784.
Fernández-Navarro, F., Hervás-Martínez, C., Gutierrez, P. A., Peña, J., & López-Granados, F. (2012). Parameter estimation of q-gaussian radial basis functions neural networks with a hybrid algorithm for binary classification. *Neurocomputing, 75*, 123–134.
Fernández-Navarro, F., Hervás-Martínez, C., Sánchez-Monedero, J., & Gutierrez, P.A. (in press). Melm-grbf: A modified version of the extreme learning machine for generalized radial basis function neural networks. *Neurocomputing*.
Guti'errez, P. A., Hervás-Martínez, C., & Martínez-Estudillo, F. J. (2011). Logistic regression by means of evolutionary radial basis function neural networks. *IEEE Transactions on Neural Networks, 22*(2), 246–263.
Howlett, R. J., & Jain, L. C. (2001). *Radial basis function networks 1: Recent developments in theory and applications. Studies in fuzziness and soft computing* (1st ed.,). Heidelberg: Physica-Verlag.
Ismail, A., & Engelbrecht, A. P. (2002). Pruning product unit neural networks. In *Proceedings of the 2002 international joint conference on neural networks* (pp. 257–262).
Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine Learning, 59*(1-2), 161–205.
Nandi, A. K., & Mämpel, D. (1995). An extension of the generalized gaussian distribution to include asymmetry. *Journal of the Franklin Institute, 332*(1), 67–75.
Sharifi, K., & Leron-Garcia, A. (1995). Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video. *IEEE Transactions on Circuits and Systems for Video Technology, 5*(1), 52–56.
Tallón-Ballesteros, A. J., & Hervás-Martínez, C. (2011). A two-stage algorithm in evolutionary product unit neural networks for classification. *Expert Systems with Applications, 38*(1), 743–754.
Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed. Data management systems). Morgan Kaufmann (Elsevier).
Yao, X. (1999). Evolving artificial neural networks. *Proceedings of the IEEE, 87*(9), 1423–1447.
Zheng, S., Li, Z., & Wang, H. . (2011). A genetic fuzzy radial basis function neural network for structural health monitoring of composite laminated beams. *Expert Systems with Applications, 38*(9), 11837–11842.