18. Salas-Morera, L., Cubero-Atienza, A.J., Ayuso-Munoz, R.: Computer-aided plant layout. Informacion Tecnologica 7(4), 39–46 (1996)
19. Sato, T., Hagiwara, M.: Idset: Interactive design system using evolutionary techniques. Computer-Aided Design 33(5), 367–377 (2001)
20. Takagi, H.: Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation. Proceedings of the IEEE 89(9), 1275–1296 (2001)
21. Tate, D.M., Smith, A.E.: Unequal area facility layout using genetic search. IIE Transactions 27, 465–472 (1995)
22. Tompkins, J., White, J., Bozer, Y., Tanchoco, J.: Facilities Planning, 4th edn. Wiley, New York (2010)
23. Tong, X.: SECOT: A Sequential Construction Technique For Facility Design. Doctoral Dissertation, University of Pittsburg (1991)
24. Wong, K.Y., Komarudin.: Solving facility layout problems using flexible bay structure representation and ant system algorithm. Expert Syst. Appl. 37(7), 5523–5527 (2010), http://dx.doi.org/10.1016/j.eswa.2009.12.080

Combining Evolutionary Generalized Radial Basis Function and Logistic Regression Methods for Classification

Adiel Castaño Méndez, Francisco Fernández-Navarro,
Pedro Antonio Gutiérrez, Manuel Baena-García,
and César Hervás-Martínez

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 of 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

Gutiérrez et al. [4] proposed a multinomial logistic regression method, combining Evolutionary Radial Basis Function (ERBF) and Logistic

Adiel Castaño Méndez
Department of Informatics, University of Pinar del Río, Cuba
e-mail: adiel@info.upr.edu.cu

Francisco Fernández-Navarro · Pedro Antonio Gutiérrez · César Hervás-Martínez
Department of Computer Science and Numerical Analysis, University of Córdoba, Spain
e-mail: i22fenaf@uco.es, pagutierrez@uco.es, chervas@uco.es

Manuel Baena-García
Departamento de Lenguajes y Ciencias de la computación, Universidad de Málaga, 29071 Málaga, Spain
e-mail: mbaena@uma.es

E. Corchado et al. (Eds.): SOCO 2011, AISC 87, pp. 263–270, 2011.
springerlink.com © Springer-Verlag Berlin Heidelberg 2011

Regression (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 Radial Basis Function (RBF) has some drawbacks, for example, its performance decreases drastically when it is applied to approximate constant valued function or when dimensionality grows. For this reason, we propose the use of a Generalized RBF (GRBF) [1], 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. In this study, we consider three main categories that can be assigned to a worker depending on the degree of permanent disability : *no disability*, *permanent disability* and *fee* (when the worker is not assigned any degree of permanent disability, but it is financially compensated).

2 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 j-th 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. 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 cuadratic 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 $c_{ji}, \tau_j, r_j \in \mathbb{R}$. Figure 1 presents the radial unit activation for the GRBF for different values of τ

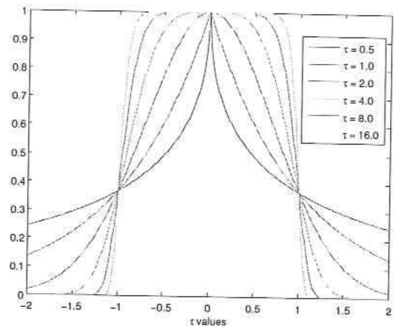

Fig. 1 Radial unit activation in one-dimensional space with $c = 0$ and $r = 1$ for the GRBF with different values of τ

3 Neuro-Logistic models

In the classification problem, some measurements x_i, $i = 1, 2, ..., k$ 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 $D = \{(\mathbf{x}_n, \mathbf{y}_n); n = 1, 2, ..., N\}$ is available, where $\mathbf{x}_n = (x_{1n}, ..., x_{kn})$ is the vector of measurements taking values in $\mathbf{\Omega} \subset \mathbb{R}^k$, and $\mathbf{y}_n$ is the class level of the n-th individual.

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

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

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

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

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_{J-1})$. 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:

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

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

4 Estimation of Neuro-Logistic Parameters

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

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

where $f_l(\mathbf{x}, \boldsymbol{\theta}_l)$ 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 [7]: 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 [5].

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

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

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

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 $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_m)$.

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 equation (6)). The specific details of this EP algorithm can be found in some previous works [6, 3].

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 $\hat{\mathbf{W}} = (\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2, ..., \hat{\mathbf{w}}_m)$, 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:

$$H : \mathbb{R}^k \rightarrow \mathbb{R}^{k+m}, (x_1, x_2, ..., x_k) \rightarrow (x_1, x_2, ..., x_k, z_1, ..., z_m), \quad (8)$$

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

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

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \tfrac{1}{N}\sum_{n=1}^{N}\left[-\sum_{l=1}^{J} y_n^{(l)}(\boldsymbol{\alpha}^l \mathbf{x}_n + \boldsymbol{\beta}^l \mathbf{z}_n) + \log\sum_{l=1}^{J}\exp(\boldsymbol{\alpha}^l \mathbf{x}_n + \boldsymbol{\beta}^l \mathbf{z}_n)\right], \quad (9)$$

where $\mathbf{x}_n = (1, x_{1n}, ..., x_{kn})$ and $\mathbf{z}_n = (z_{1n}, ..., z_{mn})$. Now, the Hessian matrix of the negative log-likelihood in the new variables $x_1, x_2, ..., x_k, z_1, ..., z_m$ is semi-definite positive. The estimated coefficient vector $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\mathbf{W}})$ 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 nonlinear 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.

5 Experiments

5.1 Database Description

The data used in this study 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. In order to apply logistic regression analyses, all nominal variables of the problems have been transformed to binary ones, resulting in a total of 51 variables. From the synthesis medical reports, we obtained the attributes Age (x_1), sex (x_2), occupation (x_{3-21}), sick leave period (x_{22}), and diseases (x_{23-42}). From the proceedings of the sessions held by the disability assessment team: Classification (permanent disability degree) (x_{43-46}), contingency (x_{47-50}) and period of time between examinations (x_{51}). Furthermore, the occupational repercussion information has been taken into account when evaluating it as low, middle or high. The classification (permanent disability degree) is grouped into: (i) No disability (ND), (ii) Permanent disability (PD), (iii) Fee (F).

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

5.2 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): (1) The k Nearest Neighbour (k-NN) classifier, adjusting the value of k using a nested 10-fold cross-validation; (2) A Gaussian Radial Basis Function Network (RBFNetwork) available in the WEKA workbench [7]; (3) Both standard logistic regression algorithms presented in Section 4: SimpleLogistic (SLogistic) and MultiLogistic (MLogistic); (4) The Naive Bayes standard learning algorithm (NaiveBayes) [7].

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 [4] 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 [4] and GRBF methodologies.

Table 1 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 [2]. 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.

Table 1 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 $\alpha = 0.05$.

	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	$\mathbf{90.70 \pm 3.02}$	**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	$\mathit{89.86 \pm 9.40}$	*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

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 1 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.

6 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 sinergy between these two techniques has been experimentally proved using a permanent disability classification problem.

Acknowledgements. 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

1. Castano, A., Hervás-Martínez, C., Gutiérrez, P.A., Fernández-Navarro, F., Garcia, M.M.: Classification by evolutionary generalized radial basis functions. In: ISDA, pp. 203–208. IEEE Computer Society, Los Alamitos (2009)
2. Corchado, E., Arroyo, Á., Tricio, V.: Soft computing models to identify typical meteorological days. Logic Journal of IGPL (2010) (in press)
3. Gutiérrez, P.A., Hervás-Martínez, C., Lozano, M.: Designing multilayer perceptrons using a guided saw-tooth evolutionary programming algorithm. Soft Computing 14(4), 599–613 (2010)
4. Gutiérrez, P.A., Hervás-Matínez, C., Martínez-Estudillo, F.J.: Logistic regression by means of evolutionary radial basis function neural networks. IEEE Transactions on Neural Networks (in press)
5. Landwehr, N., Hall, M., Frank, E.: Logistic model trees. Machine Learning 59(1-2), 161–205 (2005)
6. Martínez-Estudillo, F.J., Hervás-Martínez, C., Gutiérrez, P.A., Martínez-Estudillo, A.C.: Evolutionary product-unit neural networks classifiers. Neurocomputing 72(1-2), 548–561 (2008)
7. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques, 2nd edn. Data Management Systems. Morgan Kaufmann, San Francisco (2005)

Short-Term Wind Energy Forecasting Using Support Vector Regression

Oliver Kramer and Fabian Gieseke

Abstract. Wind energy prediction has an important part to play in a smart energy grid for load balancing and capacity planning. In this paper we explore, if wind measurements based on the existing infrastructure of windmills in neighbored wind parks can be learned with a soft computing approach for wind energy prediction in the ten-minute to six-hour range. For this sake we employ Support Vector Regression (SVR) for time series forecasting, and run experimental analyses on real-world wind data from the NREL western wind resource dataset. In the experimental part of the paper we concentrate on loss function parameterization of SVR. We try to answer how far ahead a reliable wind forecast is possible, and how much information from the past is necessary. We demonstrate the capabilities of SVR-based wind energy forecast on the micro-scale level of one wind grid point, and on the larger scale of a whole wind park.

1 Introduction

Wind energy forecasting is an important aspect for balancing authorities in a smart grid. Up to now, the integration of decentralized energy into the grid is as good as ignored. It is estimated that the stability of the energy grid decreases, if the amount of ignored renewable energy exceeds about 15% to 20%. But wind resources are steadily increasing. For a reasonable integration of volatile resources like wind, a precise prediction for subhourly scheduling becomes necessary. Precise forecast will allow balancing and integrating of multiple volatile power sources at all levels of the transmission and distribution grid [10]. Soft computing can play an important role

Oliver Kramer
ICSI Berkeley, 1947 Center St, Berkeley CA 94704, USA
e-mail: okramer@icsi.berkeley.edu

Fabian Gieseke
TU Dortmund, Otto-Hahn-Str. 14, 44227 Dortmund, Germany
e-mail: fabian.gieseke@tu-dortmund.de

E. Corchado et al. (Eds.): SOCO 2011, AISC 87, pp. 271–280, 2011.
springerlink.com © Springer-Verlag Berlin Heidelberg 2011