Expert Systems with Applications 39 (2012) 8350–8355

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, 2011), 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: