Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 195 (2009) 543–551
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

O.R. Applications

Logistic evolutionary product-unit neural networks: Innovation capacity of poor Guatemalan households

Carlos R. García-Alonso [a,*], Jorge Guardiola [b], César Hervás-Martínez [c]

[a] *Department of Management and Quantitative Methods (ETEA) University of Cordoba, Escritor Castilla Aguayo 14004 Cordoba, Spain*
[b] *Department of Applied Economics, University of Granada, Spain*
[c] *Department of Computing and Numerical Analysis of the University of Córdoba, Spain*

Received 16 November 2006; accepted 8 February 2008
Available online 20 February 2008

Abstract

A new logistic regression algorithm based on evolutionary product-unit (PU) neural networks is used in this paper to determine the assets that influence the decision of poor households with respect to the cultivation of non-traditional crops (NTC) in the Guatemalan Highlands. In order to evaluate high-order covariate interactions, PUs were considered to be independent variables in product-unit neural networks (PUNN) analysing two different models either including the initial covariates (logistic regression by the product-unit and initial covariate model) or not (logistic regression by the product-unit model). Our results were compared with those obtained using a standard logistic regression model and allow us to interpret the most relevant household assets and their complex interactions when adopting NTC, in order to aid in the design of rural policies.

Keywords: Neural networks; Logistic regression; Product-unit; Evolutionary algorithms; Sustainability; Poor households

1. Introduction

The logistic model, as a nonlinear regression model, is a special case of generalized linear model methodology (McCullagh and Nelder, 1989) where the assumptions of normality and constant variance of residuals are not satisfied. Logistic regression (LR) models have demonstrated their accuracy in many classification frameworks (Bose and Pal, 2006; Cook et al., 2006; De Andrés et al., 2006; Kiang, 2003), often yielding easy to interpret classifiers that decision makers can use to select the adequate model in real supervised learning situations. In binary problems, where the goal is to distinguish the appropriate class ($Y = 0$ or $Y = 1$) of every observation in a set by using k observed predictor variables (input variables or covariates) $x_1, x_2, \ldots, x_k$, LR has demonstrated its scientific potential for predicting the response variable. However, when the stringent assumptions of additive and pure linear effects of the covariates cannot be assumed, classical LR can be unstable – for example, multicollinearity problems (Hosmer and Lemeshow, 1989; Friedman et al., 2000). Numerous methods can be used to overcome this problem, such as sigmoidal feed-forward neural networks, projection pursuit learning, generalized additive models, and multivariate adaptive splines (Hervás-Martínez and Martínez-Estudillo, 2007).

In this paper, LR is improved (to include the nonlinear effects of the covariates) taking the hybridation of linear and product-unit models into account. Product-units (PU) are nonlinear basis functions designed using the product of the covariates raised to arbitrary powers (real values). PU can be considered as independent variables in product-unit neural networks (PUNN) to express strong covariate interactions (Durbin and Rumelhart, 1989; Ismail and Engelbrecht, 2000; Martínez-Estudillo et al., 2006a,b). In this way, the LR model can be structured,

* Corresponding author. Tel.: +34 957 222 168; fax: +34 957 222 101.
E-mail addresses: cgarcia@etea.com (C.R. García-Alonso), jguardiola@ugr.es (J. Guardiola), chervas@uco.es (C. Hervás-Martínez).

0377-2217/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2008.02.013

on the one hand, with only PU: logistic regression by the product-unit model (LRPU) or, on the other hand, with both PU and the initial covariates: logistic regression by the product-units and initial covariates model (LRLPU). These two new approaches can improve other classification methods when covariate interactions are expected to be especially relevant.

The estimation of the coefficients of LRPU and LRLPU models was carried out in two sequential steps (Hervás-Martínez and Martínez-Estudillo, 2007). In the first step, an evolutionary algorithm (EA) was applied to both design the structure (number of PU) and train the weights (PU exponents) in the PUNN. Due to the size and complexity of the search space, the EAs cannot be used to estimate all the LRPU and LRLPU coefficients. In this respect, some experiments have been carried out but the results were not satisfactory enough as also happened in the literature (Houck et al., 1996; Michalewicz and Schoenauer, 1996; Houck et al., 1997) because EAs are not suitable tools for optimum search. On the other hand, EAs can easily detect local optima in the search space (sometimes they can also locate the global optimum, if it exists) but their convergence to the problem global optima is usually too slow. Therefore, EAs can quickly find good solutions but they need many generations to reach the optimum solution (Joines and Kay, 2002). Due to these facts, our second step is justified by the need to improve EA precision. A local optimization algorithm (standard maximum likelihood method) was applied with the specific purpose of fitting the structure of the model, once the number of PU – basis functions – was determined by the EA. Finally, a backward method was used to prune correlated and non-significant covariates.

Production and commercialisation of non-traditional crops (NTC) in Guatemala were promoted during the 70s and the 80s by the international development organizations as well as the Governments as a strategy to reduce the poverty of adopters (Carletto et al., 1999; Goldín, 2003; Hamilton and Fisher, 2005; MacFarlane, 1996). This initiative allowed many rural households to escape the poverty trap and it was motivated by two circumstances. First, very few physical assets, which however need heavy investment, were enough for rural households to be able to cultivate these crops. Second, the crops have short growth cycles and therefore can produce returns of capital in the short term (Damiani, 2000).

In Guatemala, traditional agriculture, based basically on corn and beans, dominates the productive structure of rural households. The commercialisation of the resulting products is limited and barely lets them obtain enough money for their own needs. Moreover, the corn is expected to be negatively affected (massive importing) by Central American Free Trade Agreement development (Monge-González et al., 2003).

The objective of this paper was 2-fold. First of all, we wanted to analyse the accuracy of our hybrid classification algorithms as an alternative for classifying sets of observations in complex environments when the relevance of covariate interactions was expected to be very relevant. Our second objective was to use the resulting classification models to determine useful patterns for the design of rural policies. To achieve these two goals, LRPU and LRLPU algorithms were compared to standard LR to determine which variables (assets) influence the decision process in the adoption of NTC in poor households in the Guatemalan Highlands.

The data used in this paper come from a field study carried out in the Guatemalan Highlands (San Marcos and Quetzaltenango). Three hundred and seventy nine high poverty rate rural households in 8 villages were surveyed and described (PMA-MAGA, 2002). This sample was classified in two groups: non-traditional and traditional households. According to Von Braun et al. (1989), a traditional household (TH) was defined as one that devoted less than 10% of its cultivated area to NTC; higher percentages meant that the household was non-traditional (NTH). In this context, a crop was considered to be non-traditional when its production was market-oriented. Based on this database, a standard 10-fold cross-validation procedure, based on geographical conglomerates, was randomly structured to analyse the classification methodologies proposed. In all partitions (training and generalization sets), the correct classification rate (CCR) and the partial classification rate (PCA) as well as the producer's accuracy (PA) and the user's accuracy (UA) were determined and analysed.

Results showed that the improved LR models, LRPU and LRLPU, almost always achieved better performance in classifying poor rural households as TH or NTH in the Guatemalan Highlands. The analysis of specific covariate interactions is especially relevant for decision makers who can discriminate their positive and negative effects in promoting sustainable processes and good practice.

This paper is structured as follows: in Section 2, our improved classification methods are described; Section 3, briefly describes the relevance of NTC in poor rural households in Guatemala; in Section 4 the experimental design is described; classification results from our hybrid models and the most relevant findings are statistically described and analysed in Section 5 and, finally, some illustrative conclusions are drawn in Section 6.

2. Classification methods

Nowadays (Landwehr et al., 2005; Wei-Yu and Dongsong, 2006), LR has been taken into account more and more by the machine learning community in general as well as by those researchers interested in artificial neural networks because the two methods are closely related.

2.1. Logistic regression (LR) with product-unit (PU) covariates

Let $\boldsymbol{X} \in R^p$ be a set which denotes the corresponding vector of covariates, that is, a set of continuous variables observed without error and let us consider n observations of such variables noted in the form $(x_{il})_{n\times p}$. Let $\boldsymbol{y} = (y_1, \ldots, y_n)'$, a random sample extracted from a Bernoulli random population variable Y, that is, a binary response variable associated with covariate observations (x_{il}).

We say that the variables $(\boldsymbol{X}, Y)$ satisfy a LR model, if

$$y_i = f(\mathbf{x}_i, \boldsymbol{\beta}) + \varepsilon_i, \quad \text{for } i = 1, \ldots, n, \tag{1}$$

where $f(\mathbf{x}_i, \boldsymbol{\beta}) = \sum_{j=0}^{p} x_{ij}\beta_j = \mathbf{x}_i'\boldsymbol{\beta}$, $\mathbf{x}_i' = (1, x_{i1}, x_{i2}, \ldots, x_{ip})$, $\boldsymbol{\beta}' = (\beta_0, \ \beta_1, \ \beta_2, \ldots, \beta_p)$ and ε_i are random variables with mean zero. A common choice for $p_i(\mathbf{x}_i, \boldsymbol{\beta})$ that maps the real line onto the unit interval $[0, 1]$ is

$$p_i(\mathbf{x}_i, \boldsymbol{\beta}) = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i'\boldsymbol{\beta})}. \tag{2}$$

This logistic response function (2) can be easily linearized by the transformation

$$\log \frac{p_i}{1 - p_i} = \mathbf{x}_i'\boldsymbol{\beta}, \tag{3}$$

where the ratio $\frac{p_i}{1-p_i}$ is the odd. Due to the nonlinearity of the model, coefficient estimation must be carried out using iterative algorithms.

In this study, we propose a new alternative (Martínez-Estudillo et al., 2006a,b) for the non-linear function $f(\mathbf{x}, \boldsymbol{\beta})$ by the inclusion of product-unit functions in its structure, establishing therein two parts: the first one is linear while the other is non-linear and made up of covariates formed as product-unit functions in the form

$$B_j = B(\mathbf{x}, \mathbf{w}_j) = \prod_{l=1}^{p} x_l^{w_{jl}} \quad j = 1, \ldots, m, \ \ l = 1, \ldots, p. \tag{4}$$

The non-linear part of the function can be represented as a PUNN model (Durbin and Rumelhart, 1989). The network has p inputs that represent the covariates of the model, m nodes in the hidden layer: the number of basis functions and one node in the output layer (in a two-class classification problem, there is only one dependent variable Y that can only have values of 0 or 1). The activation function of the jth node of the hidden layer is given by (4), where w_{jl} is the weight of the connection between the input node l and the hidden node j. The activation of the node in the output layer is given by $\sum_{j=1}^{m}\beta_j B_j$, and the transfer function of a hidden node is the identity. In this way, a LR by product-units and initial covariates model, LRLPU, is given by

$$f(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{x}'\boldsymbol{\alpha} + \mathbf{B}'(\mathbf{x}, \mathbf{W})\boldsymbol{\beta}, \tag{5}$$

where $\mathbf{x}' = (1, x_1, \ldots, x_p)$, $\mathbf{B}'(\mathbf{x}, \mathbf{W}) = [B_1(\mathbf{x}, \mathbf{w}_1), \ldots, B_m(\mathbf{x}, \mathbf{w}_m)]$, $B_j(\mathbf{x}, \mathbf{w}_j)$ being (4), and the parameters $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \beta, \mathbf{W})$, $\boldsymbol{\alpha}' = (\alpha_0, \alpha_1, \ldots, \alpha_p)$, $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_m)$ and $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$ being $\mathbf{w}_j' = (w_{j1}, \ldots, w_{jp})$ in which $w_{jl} \in R$. The LRPU model only includes $\mathbf{B}'(\mathbf{x}, \mathbf{W})\boldsymbol{\beta}$ (product-units).

So the new conditional distribution is

$$p(\mathbf{x}, \boldsymbol{\theta}) = \frac{\exp(\mathbf{x}'\boldsymbol{\alpha} + \mathbf{B}'(\mathbf{x}, \mathbf{W})\boldsymbol{\beta})}{1 + \exp(\mathbf{x}'\boldsymbol{\alpha} + \mathbf{B}'(\mathbf{x}, \mathbf{W})\boldsymbol{\beta})}. \tag{6}$$

In this case, the decision boundaries are the generalized surface response models (Myers and Montgomery, 2002). If we have a training data set $D\{(\mathbf{x}_i, y_i)\}$ for $i = 1, \ldots, n$, where $\mathbf{x}_i > 0 \ \forall \ i$, we will use a maximum likelihood method to estimate parameters α and β – in a second step – because $\mathbf{W}$ (in the linear predictor $\mathbf{x}_i'\boldsymbol{\alpha} + \mathbf{B}'(\mathbf{x}_i, \mathbf{W})\boldsymbol{\beta}$) was estimated previously by an evolutionary algorithm (EA). Each sample observation follows a Bernoulli distribution, so since the observations are independent, the likelihood function is just

$$L(y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} p_i^{y_i}(1 - p_i)^{1-y_i}. \tag{7}$$

And the negative log-likelihood for these observations is

$$\ln L(y_1, y_2, \ldots, y_n, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^{n} [y_i f(x_i, \boldsymbol{\alpha}, \boldsymbol{\beta}) - \ln(1 + e^{f(x_i, \boldsymbol{\alpha}, \boldsymbol{\beta})})]. \tag{8}$$

Numerical search methods could be used to compute the maximum likelihood estimates (MLE) $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$. However, it turns out that we can use iteratively reweighed least squares (IRLS) to actually find the MLE. We use the SPSS© computer program that implements IRLS for the LR model. In order to define the LR using only product-units as covariates, the LRPU model simplifies Eq. (5) establishing $\boldsymbol{\alpha} = (\alpha_0, 0, \ldots, 0)$; in this form, we obtain LR models where the linear and non-linear structure of the $f(\mathbf{x}, \boldsymbol{\beta})$ function has been modelled only with the associated covariates to underlying interactions within the initial covariates.

2.2. The estimation of LRPU and LRLPU coefficients

The methodology proposed to estimate both LRPU and LRLPU parameters is a two-step procedure based on the combination of an EA (global explorer) and a local optimization procedure (local exploiters) carried out by a standard maximum likelihood optimization method. In the first step, the EA is applied to design the structure and train the weights of the PU neural network. The population-based EA for architectural design and estimation of real-parameters has points in common with other EAs in the bibliography (Angeline et al., 1994; Yao and Liu, 1997; García-Pedrajas et al., 2002). It begins the search with an initial population, and for each generation the population is updated using a population-update algorithm. The evolutionary process determines the number m of potential basis functions of the model and the corresponding vectors $\mathbf{w}_j$ of exponents in (4).

We apply a population-based EA for architectural design and the estimation of weights in the PUNN. The algorithm begins the search with an initial population, and on each generation the population is updated. It uses the replication operator and two types of mutation operators: structural and parametric. The structural mutation implies a modification of the structure of the function performed by the network and allows an exploration of different regions of the search space. The parametric mutation modifies the coefficients of the model using a self-adaptive annealing algorithm. Crossover is not used due to its disadvantages in evolving artificial neural networks (Angeline et al., 1994).

The algorithm (more details in Martínez-Estudillo et al., 2006a) begins with the random generation of a larger number of networks than the amount of networks used during the evolutionary process. We generate $10N$ networks randomly and then we select the best N . Next, we construct the base initial population of size N and evaluate the fitness score for each individual in the population using the objective function. Then, the algorithm copies the best individual to the next generation and the best 10% of population substitutes the worst 10% of the individuals. Over this intermediate population, we apply parametric mutation operators to the best 10% of population and structural mutation to the rest of the population.

At this form, the weight vector $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m)$ is estimated by means of an evolutionary neural network algorithm that optimizes the error function given for a g model, in this case, by the log-likelihood function (n observations)

$$L(\boldsymbol{\beta}, \mathbf{W}) = \sum_{l=1}^{n} \left\{ y_l f(\mathbf{x}_l, \boldsymbol{\beta}, \mathbf{W}) - \log\left(1 + e^{f(\mathbf{x}_l, \boldsymbol{\beta}, \mathbf{W})}\right) \right\}. \tag{9}$$

The fitness measure is a strictly decreasing transformation of the error function $L(\boldsymbol{\beta}, \mathbf{W})$ given by

$$A(g) = \frac{1}{1 + L(\boldsymbol{\beta}, \mathbf{W})}, \quad \text{where } 0 < A(g) \leqslant 1. \tag{10}$$

Parametric mutation is accomplished for each coefficient w_{jl} and β_j of the model with Gaussian noise

$$w_{jl}(t+1) = w_{jl}(t) + \xi_1(t), \quad \beta_j(t+1) = \beta_j(t) + \xi_2(t), \tag{11}$$

where $\xi_k(t) \in N(0, \alpha_k(t))$, $k = 1, 2$, represents a one-dimensional normally distributed random variable with mean 0 and variance $\alpha_k(t)$. Once the mutation is performed, the fitness of each individual is recalculated and the usual simulated annealing (Kirkpatric et al., 1983) is applied. Thus, if ΔA is the difference in the fitness function before and after the random step, the criterion is if $\Delta A \geqslant 0$ the step is accepted, and if $\Delta A < 0$ the step is accepted with a probability $\exp(\Delta A / T(g))$, where the temperature $T(g)$ of an individual g model is given by $T(g) = 1 - A(g)$, $0 \leqslant T(g) < 1$.

The variance $\alpha_k(t)$ is updated throughout the evolution. There are different methods to update the variance. We use the 1/5 success rule of Rechenberg (1975) that is one of the simplest but effective methods.

There are five different structural mutations that are applied sequentially to each network: node addition, node deletion, connection addition, connection deletion and node fusion; the first four followed the works of Angeline et al. (1994) and in the fifth two randomly selected nodes, a and b, are replaced by a new node c, which is a combination of the two. The connections that are common to both nodes are kept, with a weight given by

$$\beta_c = \beta_a + \beta_b, \quad w_{jc} = \frac{w_{ja} + w_{jb}}{2}. \tag{12}$$

The connections that are not shared by the nodes are inherited by c with probability 0.5 and their weights remain unchanged.

The number of hidden nodes added is calculated as $\Delta_{\mathrm{MIN}} + uT(g)[\Delta_{\mathrm{MAX}} - \Delta_{\mathrm{MIN}}]$, u being a random uniform variable in the interval $[0, 1]$, $T(g) = 1 - A(g)$ the temperature of the g neural net model, and Δ_{MIN} and Δ_{MAX} the minimum and maximum number of hidden nodes to be added. However, the connection addition and deletion mutations are performed in a slightly different way. For each mutated neural net, we apply connection mutations sequentially, first adding (or deleting) $1 + u[\Delta_{\mathrm{O}} n_{\mathrm{O}}]$ connections from the hidden layer to the output layer and then adding (or deleting) $1 + u[\Delta_{\mathrm{H}} n_{\mathrm{H}}]$ connections from the input layer to the hidden layer, u being a random uniform variable in the interval $[0, 1]$, Δ_{O} and Δ_{H} a previously defined ratio of the number of connections of both the hidden and the output layers and n_{O} and n_{H} the current number of connections in the output and the hidden layers.

Following this proposal, in our present paper we have used the following algorithm parameters: the exponents w_{ji} are randomly initialized in the interval $(-5, 5)$ and the coefficients β_{kj} are initialized in $(-5, 5)$. In addition, the maximum number of nodes in the hidden layer is $m = 4$. The size of the population is $N = 1000$. The number of nodes that can be added or removed in a structural mutation is within the $[1, 2]$ interval, and the ratio of the number of connections of the hidden and the output layers is $\Delta_{\mathrm{O}} = 0.05$ and $\Delta_{\mathrm{H}} = 0.3$.

The stopping criterion is reached whenever one of the following two conditions is fulfilled: (i) for 20 generations, there is no improvement either in the average performance of the best 20% of the population or in the fitness of the best individual; or (ii) the algorithm achieves 150 generations. We have done a simple linear rescaling of the input variables in the interval $[0.1, 0.9]$, X_i^* being the transformed variables. The lower bound is chosen to avoid input values near 0 that could produce very large values of the outputs for negative exponents. The upper bound is chosen to avoid dramatic changes in the outputs of the network when there are weights with large values (especially in the exponents).

Finally in this first step, the basis functions $B_1(\mathbf{x}, \hat{\mathbf{w}}_1), B_2(\mathbf{x}, \hat{\mathbf{w}}_2), \ldots, B_m(\mathbf{x}, \hat{\mathbf{w}}_m)$ of the best PUNN model obtained by the EA in the last generation (global search) are included in the covariate space of the LR model. We remark that the $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)$ parameters in the best PUNN model are not considered because they will be estimated at the same time as the $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_p)$ parameters (local search) using the maximum likelihood method in the second step.

In the second step, we consider a transformation of the input space adding the nonlinear transformations of the input variables given by the basis functions obtained by the EA. The model is linear in these new variables together with the initial covariates. The remaining coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are calculated by the maximum likelihood optimization method using a IRLS algorithm.

In order to select the final model, we use a backward stepwise procedure pruning successively variables sequentially to the model until further prunes do not improve the fit. At each step, we deleted the least significant ($\alpha = 0.05$) covariate to predict the response variable, that is, the one which shows the greatest critical value (p-value) in the hypothesis test, where the associated coefficient equal to zero is the hypothesis to be contrasted. The procedure finishes when all tests provide p-values smaller than the fixed significance level and the model selected fits well.

3. Influence of non-traditional crops (NTC) in poor rural households

According to Von Braun et al. (1989), Carletto et al. (1999) and Hamilton and Fisher (2003, 2005), the adoption of NTC improved a secure household food supply as a result of generating higher family income by the product commercialisation. In addition, NTC propelled the creation of both direct and indirect employment for activity supporting services, and the multiplicative effect of the generated money may be decisive for rural development and in maintaining the rural population stable. These circumstances made the adoption of NTC convenient for the development strategies.

The study of Immink and Alarcón (1993) in the Guatemalan Highlands showed that the NTC adoption increases household income. However, the authors argued the positive effect of NTC cultivation on the secure food supply of the rural population because the income increase was not associated with better household food intake scores. Some important reasons for this phenomenon were associated with previous loan interest payments and the need for capital to restructure farms.

This study analyses what covariates (household characteristics) can be considered essential in NTC adoption by mathematically determining their interactions in order to transcend inexact linear approaches.

4. Experimental design

Data used in this study come from the field work carried out in San Marcos and Quetzaltenango departments in the Guatemalan Highlands. In both departments, the PMA-MAGA (2002) classification characterized the majority of the rural households with high poverty rates. Nevertheless, this fact contrasts with some successful experiences in adopting, producing and the commercialisation NTC (Goldín, 2003). Compared to the rest of the departments in the Guatemalan Highlands, San Marcos and Quetzaltenango have got a better access by road but, on the other hand, also run a greater risk of weather disasters, mainly frosts.

Data include 379 observations (surveyed households in 2005) from 8 different villages located in four different municipalities. The selection of the households was made by simple random sampling. Villages with more than 75% urban households were previously rejected. Based on the maps of the selected village, groups of 6 households were identified and numbered. These groups were finally used to randomly select the final sample.

Surveyed households were classified into two classes (Von Braun et al., 1989; Immink and Alarcón, 1993): those that had less than 10% of their cultivated area devoted to NTC ($Y = 1$ for classification purposes. Traditional Households, TH) – 246 in total, 64.91%-the rest being non-traditional households NTH – 133 households, 35.09% ($Y = 0$).

Table 1
Analysed variables in the 379 household sample (Guatemalan Highlands): 246 traditional households (TH) and 133 non-traditional households (NTH)

Variable and description	Mean or percentage	
	TH	NTH
ED: household head education level: (1) none; (2) basic; (3) high school level and (4) higher than high school level	Mean: 1.3	Mean: 1.2
AG: age of the household head (years)	Mean: 46.1	Mean: 44.1
SE: household head sex (male or female)	91.7% males	87.8% males
WH: quotient between the weekly worked hours on the household farm and the number of household members that is devoted to this activity (hours per week/household member)	Mean: 2.1	Mean: 3.1
CA: farm cultivated area (ha)	Mean: 0.644	Mean: 0.497
IR: does the household farm have an irrigation system? (yes or no)	14.2% with irrigation	39.1% with irrigation
SA: quotient between the total (sum) household salaries and the total number of the household members (US $/member)	Mean: 0.9867	Mean: 0.9733
TR[a]: is the household considered as a traditional one? (yes or no)	85% traditional	69.2% traditional
RE: does the household receive remittances from abroad in 2004? (yes or no)	23.6% receives	15.8% receives
FM: family members	Mean: 5.8	Mean: 6.1

[a] Household attitude to the corn cultivation. It is a survey-based variable based on the answers (“is a tradition” or “for eating”) to the question: Why do you cultivate corn?

For the classification analysis, a 10-fold cross-validation procedure was chosen (Goutte, 1997). Folds, however, were randomly designed based on geographical conglomerates defined for every surveyed village. This structure guaranteed the spatial representativity of the results. The covariates (Table 1) were selected and surveyed according to previous studies like Von Braun et al. (1989) and Immink and Alarcón (1993).

A previous correlation analysis showed significant p-values at 0.05 between AG and WH (0.000), AG and CA (0.005), WH and SA (0.000), WH and FM (0.000) and, finally, CA and FM (0.018). Taking these results into account, strong relationships between covariates were expected and LRPU and LRLPU models were especially adequate to detect them. Correlated variables were included in the analysis because they are not a serious problem when the purpose is prediction (Judge et al., 1982; Torres et al., 2005).

In order to evaluate and compare the accuracy of the proposed classification models, correct classification rate (CCR), partial classification rate (PCR), producer's accuracy (PA) as well as user's accuracy (UA) ratios were calculated for both the training and generalization sets (Borghys and Yvinec, 2006). The first one (CCR) can be defined as the percentage of the total correct classified observations to the total number of observations. PCR is equivalent to CCR for each target response ($Y = 1$ and $Y = 0$).

PA and UA were calculated for the LRPU and LRLPU best models. The PA is the number of observations correctly classified as a given class to the total number of observations that belongs to this class. It can be interpreted as the probability of a correct classification. UA is calculated as the number of observations correctly classified in a class to the total number of observations that was classified as this class by the algorithm. It is a concept to avoid any probability of false alarms. The best classification method – perfect – is that where both PA and UA parameters are equal to one (Borghys and Yvinec, 2006).

5. Results

In the analysis of the generalization sets for the 10-folds, LRPU and LRLPU models showed the best CCR_G global results, Table 2. Only in two partitions (#4, the worst, and #8), LR partially dominated our proposed models. In 6 of the 10-folds the LRPU model demonstrated that the interactions are more relevant than the linear part of the equation (Table 2) because both LR (linear part) and LRLPU (linear part and interactions) showed worst mean CCR_G results. The standard deviation of CCR_G scores for LRPU is a little bit higher than those obtained in LR and LRLPU, so the linear part of the equation tends to stabilise variability on analysing all the partitions. The reason for the higher LRPU standard deviation is partition #4 where CCR_G was low compared to the remaining 9-folds.

The best classification model was obtained in partition #7 (Table 3). LRPU and LRLPU models showed, in this

Table 2
CCR for the training and generalization sets for LR, LRPU and LRLPU models in a 10-fold experimental design based on geographical conglomerates

Partition	CCR_T			CCR_G		
	LR	LRPU	LRLPU	LR	LRPU	LRLPU
1	72.7	**77.1**	75.7	65.8	**71.1**	68.4
2	74.2	**77.1**	76.5	65.8	65.8	**68.4**
3	74.8	76.0	**78.0**	65.8	**68.4**	**68.4**
4	75.4	76.5	**77.4**	**63.2**	55.3	57.9
5	73.3	**75.1**	**75.1**	**73.7**	**73.7**	71.1
6	75.7	75.4	**79.2**	65.8	68.4	**76.3**
7	72.1	**76.0**	75.1	84.2	**89.5**	86.8
8	72.4	74.5	**75.4**	**76.3**	71.1	68.4
9	72.7	74.8	**75.1**	78.9	**83.8**	81.6
10	72.4	74.8	**75.7**	73.7	**81.6**	78.9
Mean	73.57	75.73	76.32	71.32	**72.87**	72.62
SD[a]	1.34	0.96	1.43	7.05	9.88	8.33

[a] Standard deviation.

Table 3
Rate of the number of cases that were classified correctly for the best (#7) and worst (#4) partitions of a 10-fold cross-validation procedure (based on geographical conglomerates) using logistic regression (LR), full nonlinear LRPU and LRLPU models (results are structured using the following schemata: LR/LRPU/LRLPU)

	Training		PCR[c] (%)	Generalization		PCR (%)
P. Resp.[a]	$Y = 1$	$Y = 0$		$Y = 1$	$Y = 0$	
T. Resp.[b]			Best (#7)			
$Y = 1$	195/199/196	25/21/24	88.6/90.5/89.1	26/26/26	0/0/0	100/100/100
$Y = 0$	70/61/61	51/60/60	42.1/49.6/49.6	6/4/5	6/8/7	50.0/66.7/58.3
CCR			72.1/**76.0**/75.1			84.2/**89.5**/86.8
T. Resp.[b]			Worst (#4)			
$Y = 1$	201/199/200	21/23/22	90.5/89.6/90.1	20/20/20	4/4/4	83.3/83.3/83.3
$Y = 0$	63/57/55	56/62/64	47.1/52.1/53.8	10/13/12	4/1/2	28.6/7.1/14.3
CCR			75.4/76.5/**77.4**			**63.2**/55.3/57.9

Household farms without non-traditional crops ($Y = 1$) and with non-traditional crops ($Y = 0$).
[a] Predicted response.
[b] Target response.
[c] PCR: partial correct rate (%).

Table 4
Best models for LR, LRPU and LRLPU using the seven 10-fold (best fold for CCR_G)

LR [% attributes = 100, # coefficients = 11, CCR_G = 84.2]											
Variables	Constant	ED[a]	AG[a]	SE[a]	WH[a]	CA[a]	IR[a]	SA[a]	TR[a]	RE[a]	FM[a]
Coefficient	−1.28	−2.97[b]	−2.22[c]	.29	3.86[b]	−4.52[c]	1.64[b]	1.46	.16	−.33	2.28[c]
Std error	.90	1.14	1.05	.54	.85	1.90	.41	1.16	.45	.43	1.04
LRPU [% attributes = 90, # coefficients = 23, CCR_G = 89.5]											
Variables	Constant	B_1	B_2	B_3	B_4	$B_1 = (SE^a)^{0.977}\,(WH^a)^{0.974}\,(CA^a)^{5.504}\,(IR^a)^{4.151}\,(SA^a)^{-1.661}\,(TR^a)^{-0.919}$					
Coefficient	−2.31[b]	22.35[c]	−32.94[b]	.58[b]	1.55[b]	$B_2 = (ED^a)^{-0.297}\,(SE^a)^{0.148}\,(WH^a)^{2.396}\,(CA^a)^{1.247}\,(TR^a)^{2.393}$					
Std error	.28	10.10	9.80	.10	.26	$B_3 = (IR^a)^{4.004}\,(TR^a)^{-0.835}$					
						$B_4 = (ED^a)^{-0.793}\,(AG^a)^{-0.7}\,(WH^a)^{1.22}\,(TR^a)^{0.401}\,(FM^a)^{0.439}$					
LRLPU [% attributes = 90, # coefficients = 29, CCR_G = 86.8]											
Variables	Constant	SE[a]	WH[a]	CA[a]	SA[a]	TR[a]	FM[a]	B_1	B_2	B_3	B_4
Coefficient	−3.75[b]	.32	2.20	−3.43	1.61	.84	1.35	28.61[b]	−34.20[b]	.62[b]	1.25[b]
Std error	1.11	.56	1.58	2.37	1.20	.76	1.21	10.83	11.70	.15	.31

[a] Standardized variables in the range [0.1, 0.9]. Signed by an * in the text.
[b] Significant for a coefficient $\alpha = 0.01$.
[c] Significant for a coefficient $\alpha = 0.05$.

case, their potential on analysing the generalization set. Both the PCR_G (considering $Y = 1$ – TH and $Y = 0$ – NTH) and the CCR_G scores demonstrated their precision, LRPU being the best. The learning capacity of LRPU and LRLPU models is really very important reaching excellent 89.5% and 86.8% CCR_G's for the generalization set (better scores than those obtained in the training set).

Taking the best LRPU model into consideration (partition #7), all the covariates were selected for it except RE* (rescaled RE, Table 1). According to these results, the propensity of a household to cultivate NTC is not influenced by the reception of remittances. The most relevant variable in the interactions calculated (Table 4) is TR* that evaluated the head of the household attitude for facing up the corn cultivation. This variable appears in all the interactions (B_1–B_4).

The analysis of LRPU best model interactions is quite complicated. Taken into account as a reference, a precarious household where its head has no education level ($ED^* = 0.1$, no studies), is young ($AG^* = 0.376$, 30 years old), is male ($SE^* = 0.9$), who owns a low-medium farm size ($CA^* = 0.329$, 1.4 ha), without an irrigation system ($IR^* = 0.1$), with a traditional point of view about corn cultivation ($TR^* = 0.9$) and, finally, with a low-medium family size ($FM^* = 0.26$, 4-members), the influence of the total household salary per family member (SA^*) is always neutral or slightly positive. On the other hand, the influence of the family working hours on the household farm per week and member (WH^*) is positive from 4.5 to 9 (the maximum) hours/(week and member) and negative from 0 to 4.5 hour/(week and member). That is, a local minimum is reached. The increase in the educational level

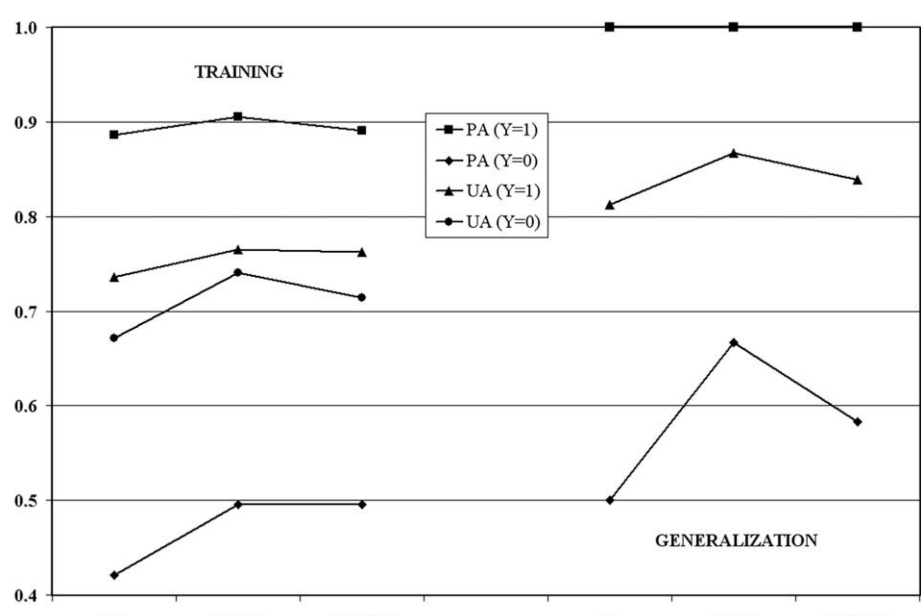

Fig. 1. Producer's accuracy (PA) and user's accuracy (UA) for the best partition (#7).

(ED^*) has a relevant and positive influence on the possibility of our previously designed head of the household of being NTH, SA^* remains neutral or, sometimes arguably, slightly positive but now the positive effect of WH starts as of 1.35 hour/(week and member).

Following our analysis based on the above-mentioned household head, the number of family members (FM^*) is a handicap for being NTH. For example, a 6-member family (37.39% of the household heads that were 30 years old had a family equal or greater than 6-members) slows down the positive effect of the WH^* to 2.25 hours/(week and member) – SA^* behaviour remains the same – when ED^* was 0.3666 (primary educational level).

Considering again a 4-member family and an educational level of $ED^* = 0.1$ (no studies), an increase on the household cultivated area is really very positive in transforming a TH into a NTH. An increase in the educational level always promotes the tendency to be NTH, but SA^* still has no significant effect.

The learning capacity of LRPU (the best) and LRLPU as well as LR is demonstrated by calculating both the PA and the UA. In the best partition (#7), PA and UA scores were greater for the generalization set than that for the training one, so the models obtained generalized quite well. LRPU again showed the best PA and UA ratios and they sometimes reached excellent values over 0.8, Fig. 1 (In the generalization set PA ($Y = 1$) and UA ($Y = 0$) were coincident). PA for $Y = 0$ (NTH) in the generalization set discriminates LRPU from LR and LRLPU confirming the evidence: the LRPU model dominates and, for our decisional framework, interactions are more relevant than individual covariates.

6. Conclusions

In this paper, we have proposed to use new improvements for the classic LR classification models based on logistic regression and product-units in a complex environment where interactions between covariates are necessary and welcome: logistic regression by the product-unit model (LRPU) and logistic regression by the product-units and initial covariates model (LRLPU). As happens in the nature, the interpretability of these interactions is complicated because they have more relevance than individual covariates. In order to make it easier, specific scenarios (fixing some covariate values) can be designed to evaluate the effect of the remaining ones.

LRPU, the best, and LRLPU models have demonstrated their adequacy, adaptability and interpretability in analysing dichotomous classification problems in a very complex environment, like in the Guatemalan Highlands. The analysis of the relationships between variables that describe the socio-productive structure of poor rural households cannot be based on the basic linear models but on those that understand complex interactions. As expected, variables with positive effects can only be considered really positive in specific zones of their range (local optimums). Due to this, it can be difficult to give conclusive statements, but we can approximate useful conclusions for designing rural development programs that should take into consideration complex, but now predictable, relationships between variables.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the Spanish Department of Research of the Ministry of Education and Science under the TIN2005-08386-C05-02 projects. FEDER also provided additional funding. Moreover, the authors would like to thank the Food and Agriculture Organization (FAO), the Ministry of Agriculture of Guatemala (MAGA), the Mesoamerican Food Security Early Warning System (MFEWS), Universidad Rafael Landívar from Guatemala.

References

Angeline, P.J., Saunders, G.M., Pollack, J.B., 1994. An evolutionary algorithm that constructs recurrent neural networks. IEEE Transactions on Neural Networks 5, 54–65.

Borghys, D., Yvinec, Y., 2006. Supervised feature-based classification of multi-channel SAR images. Pattern Recognition Letters 27 (4), 252–258.

Bose, I., Pal, R., 2006. Predicting the survival or failure of click-and-mortar corporations: A knowledge discovery approach. European Journal of Operations Research 174 (2), 959–982.

Carletto, C., de Janvry, A., Sadoulet, E., 1999. Sustainability in the diffusion of innovation: Smallholder non-traditional agro-exports in Guatemala. Economic Development and Cultural Change 47 (2), 345–369.

Cook, D.F., Zobel, C.W., Wolfe, M.L., 2006. Environmental statistical process control using an augmented neural network classification approach. European Journal of Operations Research 174 (3), 1631–1642.

Damiani, O., 2000. The state and nontraditional agricultural exports in Latin America: Results and lessons of three case studies. In: Working Paper prepared for the Conference on Development of the Rural Economy and Poverty Reduction in Latin America and the Caribbean, March 24, New Orleans.

De Andrés, J., Landajo, M., Lorca, P., 2006. Forecasting business profitability by using classification techniques: A comparative analysis based on a Spanish case. European Journal of Operations Research 167 (2), 518–542.

Durbin, R., Rumelhart, D., 1989. Product-units: A computationally powerful and biologically plausible extensión to backpropagation networks. Neural Computation 1, 133–142.

Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: A statistical view of boosting. The Annals of Statistics 38 (2), 337–374.

García-Pedrajas, N., Hervás-Martínez, C., Muñoz-Pérez, J., 2002. Multiobjetive cooperative coevolution of artificial neural networks. Neural Networks 15 (10), 1255–1274.

Goldín, L., 2003. Procesos globales en el campo de Guatemala: Opciones económicas y transformaciones ideológicas, FLACSO, Guatemala.

Goutte, C., 1997. Note on free lunches and cross-validation. Neural Computation 9, 1211–1215.

Hamilton, S., Fisher, E.F., 2003. Non-traditional agricultural exports in Highland Guatemala: Understandings of risk and perceptions of change. Latin American Research Review 38 (3), 82–110.

Hamilton, S., Fisher, E.F., 2005. Maya farmers and export agriculture in Highland Guatemala: Implications for development and labor relations. Latin American Perspectives 32 (5), 33–58.

Hervás-Martínez, C., Martínez-Estudillo, F., 2007. Logistic regression using covariates obtained by product-unit neural network models. Pattern Recognition 40 (1), 52–64.

Hosmer, D.W., Lemeshow, S., 1989. Applied Logistic Regression. John Wiley & Sons, New York.

Houck, C.R., Joines, J.A., Kay, M.G., 1996. Comparison of genetic algorithm, random start, and two-opt switching for solving large location-allocation problems. Computers Operation Research 23 (6), 587–596.

Houck, C.R., Joines, J.A., Kay, M.G., 1997. Empirical investigation of the benefits of partial Lamarckianism. Evolutionary Computation 5 (1), 31–60.

Immink, M., Alarcón, J.A., 1993. Household income, food availability, and commercial crop production by smallholder farmers in the Western Highlands of Guatemala. Economic Development and Cultural Change 4 (1), 319–343.

Ismail, A., Engelbrecht, A.P., 2000. Global optimization algorithms for training product-unit neural networks. In: IJCNN, vol. 1. IEEE Computer Society, Los Alamitos CA, pp. 132–137.

Joines, J.A., Kay, M.G., 2002. Utilizing hybrid genetic algorithms. In: Sarker, R., Mahamurdian, M., Yao, X. (Eds.), Evolutionary Optimization. Kluwer Academic Publisher.

Judge, G., Hill, C., Griffiths, W., 1982. Introduction to the Theory and Practice of Econometrics. John Wiley & Sons, New York.

Kiang, M.Y., 2003. A comparative assessment of classification methods. Decision Support Systems 35 (4), 441–454.

Kirkpatric, S., Gellat, C.D.J., Vecchi, M.P., 1983. Optimization by simulated annealing. Science 220, 671–680.

Landwehr, N., Hall, M., Eibe, F., 2005. Logistic model trees. Machine Learning 59, 161–205.

MacFarlane, R., 1996. Modelling the interaction of economic and socio-behavioural factors in the prediction of farm adjustment. Journal of Rural Studies 12 (4), 365–374.

Martínez-Estudillo, A.C., Martínez-Estudillo, F., Hervás-Martínez, C., García-Pedrajas, N., 2006a. Evolutionary product-unit based neural networks for regression. Neural Networks 19, 477–486.

Martínez-Estudillo, F., Hervás-Martínez, C., Martínez-Estudillo, A.C., García-Pedrajas, N., 2006b. Hybridation of evolutionary algorithms and local search by means of a clustering method. IEEE Transaction on Systems, Man and Cybernetics, Part B: Cybernetics 36 (3), 534–546.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Chapman and Hall, Norwell MA.

Michalewicz, Z., Schoenauer, M., 1996. Evolutionary algorithms for constrained parameter optimization problems. Evolutionary Computation 4 (1), 1–32.

Monge-González, R., Loria-Sagot, M., González-Vega, C., 2003. Retos y Oportunidades para los Sectores Agropecuario y Agroindustrial de Centro América ante un tratado de Libre Comercio con los Estados Unidos. World Bank, Washington, DC.

Myers, R.H., Montgomery, D.C., 2002. Response Surface Methodology: Process and Product Optimization using Designed Experiments. Wiley, New York.

PMA-MAGA, 2002. Cartografía y análisis de la vulnerabilidad de la inseguridad alimentaria en Guatemala, PMA-MAGA, Ciudad de Guatemala.

Rechenberg, I., 1975. Evolutionstrategie: Optimierung technischer Systeme nach Prinzipien der Biologischen Evolution. Stuttgart Framman-Holzboog Verlag.

Torres, M., Hervás-Martínez, C., Amador, F., 2005. Approximating the sheep milk production curve through the use of artificial networks and genetic algorithms. Computers and Operations Research 32, 2653–2670.

Von Braun, J., Hotchkiss, D., Immink, M., 1989. Nontraditional Export Crops in Guatemala: Effects on Production, Income, and Nutrition, IFPRI Research Report 73, Washington, DC.

Wei-Yu, C., Dongsong, L.Z., 2006. Predicting and explaining patronage behaviour toward WEB and traditional stores using neural networks: A comparative analysis with logistic regression. Decision Support Systems 41 (2), 514–531.

Yao, X., Liu, Y., 1997. A new evolutionary system for evolving artificial neural networks. IEEE Transactions on Neural Networks 8 (3), 694–713.