Feature Selection for Hybrid Neuro-Logistic Regression applied to Classification of Remote Sensed Data

P.A. Gutiérrez, J.C. Fernández, C. Hervás
Department of Computer Science and Numerical Analysis, University of Córdoba
14071 Córdoba, Spain
i02gupep@uco.es, fernandezcaballero@gmail.com, chervas@uco.es

F. López-Granados, M. Jurado-Expósito and J.M. Peña-Barragán
Institute for Sustainable Agriculture, CSIC
14080, Córdoba, Spain
flgranados@ias.csic.es, montse.jurado@ias.csic.es, pa2pebaj@uco.es

Abstract

Logistic Regression (LR) has become a widely used and accepted method to analyse binary or multiclass outcome variables, since it is a flexible tool that can predict the probability for the state of a dichotomous variable. A recently proposed LR method is based on the hybridisation of a linear model and Evolutionary Product-Unit Neural Network (EPUNN) models for binary classification. This produces a high number of coefficients, so two different methods for reducing the number of initial or PU covariates are proposed in this paper, both based on the Wald test. The first method is a two-step Backward Search (BS) method and the second is based on the standard Simulated Annealing (SA) heuristic. In this study, we used aerial imagery taken in mid-May to evaluate the potential of two different combinations of LR and EPUNN (LR using PUs (LRPU), as well as LR using Initial covariates and PUs (LRIPU)) and the two proposed methods for selecting variables in the final models (BS and SA) for discriminating Ridolfia segetum patches (one of the most dominant, competitive and persistent weed in sunflower crops) in one naturally infested field of southern Spain. Then, we compared the performance of these methods to six recent classification models, our proposals obtaining a competitive performance and a lower number of coefficients.

1. Introduction

Classification problems attempt to solve the task of deciding the class membership y of an unknown data item $\mathbf{x}$ based on a data set $D = \{(\mathbf{x}_i, y_i)\}\ i = 1, ..., n$ of data items $\mathbf{x}_i$ with known class membership. The $\mathbf{x}_i$ are usually k-dimensional feature vectors, whose components are called covariates or independent variables. In most problem domains, there is no functional relationship between y and $\mathbf{x}$. In this case the relationship has to be described more generally by a probability distribution $P(\mathbf{x}; y)$; one then assumes that the data set D contains independent samples from P. From statistical decision theory, it is well known that the optimal class membership decision is to choose the class label y that maximises posteriori distribution $P(y/\mathbf{x})$. Therefore there are different approaches to data classification: one which considers only one distinction between the classes previously defined and assigns a class label to an unknown data item, and another which attempts to model $P(y/\mathbf{x})$. This latter attempt yields not only a class label for a data item, but also a probability of class membership. The most prominent representatives of the first classification approach are support vector machines. Logistic Regression (LR), artificial neural networks, k nearest neighbours (kNN), and decision trees are all members of the second class, although they vary considerably in building an approximation to $P(y/\mathbf{x})$ from data. However, in spite of the great number of techniques developed to solve classification problems, there is no optimum methodology or technique to solve specific problems. This point has encouraged the comparison and combination of different types of classification [15].

Feature selection is a key task in remote sensing data processing, particularly in the case of classification from hyperspectral images. A LR model may be used to predict the probabilities of classes based on input features, after ranking them according to their relative importance. In this work, the LR model is applied to both the feature selec-

978-0-7695-3326-1/08 $25.00 © 2008 IEEE
DOI 10.1109/HIS.2008.34

tion and the classification of remotely sensed image pixels, where more informative classification probabilities are produced naturally.

A recently proposed LR method [8] is based on the hybridisation of a linear model and Evolutionary Product-Unit Neural Network (EPUNN) models for binary classification. The estimation of the model coefficients is carried out in two phases. First, the number of PU basis functions and the exponents' vector are determined by means of an evolutionary neural network algorithm. Secondly, a standard maximum likelihood optimization method determines the rest of the coefficients in the new space given by the initial variables and the PU basis functions previously estimated. This model allows the generation of non-linear classification surfaces and the identification of possible strong interactions that can exist between the covariates that define the classification problem. These models are less complex (number of new covariates or number of exponents in these covariates) than the alternative higher order polynomial models. However, the models result in a high number of coefficients, so two different methods for reducing the number of initial or PU covariates are proposed in this paper, both based on the Wald test. The first method is a two-step Backward Search (BS) method and the second is based on the standard Simulated Annealing (SA) heuristic.

Sunflower (*Helianthus annuus* L.) is one of the most abundant crops in Andalusia, Southern Spain, with more than 320,000 ha sown annually [1]. Sunflower sowing and harvesting times are February-March and July-August, respectively, mainly grown under dry land conditions. *Ridolfia segetum* Moris (corn caraway) is a very frequent annual, umbelliferous weed that is abundant in clay soils in Andalusia. Its life cycle coincides with that of the sunflower, which enhances its competitive ability and results in an average crop yield reduction of about 32% when infestation is two *R. segetum* plants m^{-2} [5]. This weed is hard to control because it is not controlled by the pre-emergence and pre-plant incorporated herbicides used in sunflower and consequently post-emergence strategies such as tillage or hand weeding are commonly used, otherwise weed obstructs the harvester due to it still has partly green steam during the sunflower harvesting. This is a serious drawback if the harvester is equipped with yield monitor as habitually happens in precision agriculture management.

Patchy distribution of broadleaf weeds in sunflower fields is well documented [11]. However, herbicides or other control strategies are not addressed to the infested zones, but are instead applied over the entire fields. The potential for overuse or application and the corresponding eco-environmental problems are evident. To overcome the possibility of minimising the impact of inappropriate control strategy, the idea of Site-Specific Weed Management (SSWM) has been developed in the context of precision agriculture [20]. A key component of SSWM is that accurate and appropriate weed maps are required to take full advantage of site-specific herbicide applications. Mapping weed patches based on ground survey techniques on field scale is time consuming, expensive and unapproachable in field areas with difficult access. Remote sensing of weed canopies may be more efficient and suitable than field surveys and the majority of studies on discriminating weeds in cultivated systems have involved discrete broadband remote sensing (multispectral sensors) [4]. Approaches based on EPUNNs have been previously applied to remotely sensed images for agricultural objectives [7].

Thus, the goal of this work is to assess the potential of two different combinations of LR and EPUNN (LR using PUs (LRPU) and LR using Initial covariates and PUs (LRIPU)) and the two proposed methods for selecting variables in the final models (BS and SA) for discriminating *R. segetum* patches in one naturally infested field. The results indicate that, with fewer restrictive assumptions, the models proposed are able to reduce the features substantially without any significant decrease in classification accuracy.

2. Binary Logistic Regression

The binary Logistic Regression (LR) technique considers the following situation: a binary outcome variable y is observed together with a vector $\mathbf{x}_i = (1, x_{i1}, x_{i2}, ..., x_{ik})$ of covariates for each of the n_{T} training samples (assuming that the vector of inputs includes the constant term 1 to accommodate the intercept). The two-class is coded via a $0/1$ response y_i. Let p be the conditional probability associated with the first class. LR [9] is a widely used statistical modeling technique in which the probability p of the dichotomous outcome event is related to a set of explanatory variables $\mathbf{x}$ in the form:

$$\mathrm{logit}(p) = \log\left(\frac{p}{1-p}\right) = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x} \qquad (1)$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_k)$ is the vector of coefficients of the model, $\boldsymbol{\beta}^{\mathrm{T}}$ is the transposed vector and the odd of the event is $p/(1-p)$. A simple calculation in (1) shows that the probability of occurrence of an event as a function of the covariates is nonlinear and is given by:

$$p(\mathbf{x}, \boldsymbol{\beta}) = \frac{e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}}}{1 + e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}}} = \frac{e^{f_{\mathrm{LR}}(\mathbf{x},\boldsymbol{\beta})}}{1 + e^{f_{\mathrm{LR}}(\mathbf{x},\boldsymbol{\beta})}}$$

Iteratively re-weighted least squares (IRLS) is a nonlinear optimization algorithm that uses a series of weighted least squares (WLS) subproblems to find LR model maximum-likelihood coefficient estimation. The implementation of IRLS applied in this work is based on that provided in [13], using the conjugate gradient method for solving the associated matricial equation.

3. Evolutionary Product Unit Neural Networks (EPUNNs)

PUNNs are an alternative to Multilayer Perceptrons (MLPs) and are based on multiplicative nodes instead of additive ones [6]. A multiplicative node is given by:

$$\prod_j(\mathbf{x}, \mathbf{w}_j) = \prod_{i=1}^{k} x_i^{w_{ji}}$$

where k is the number of inputs and $\mathbf{w}_j = (w_{j1}, w_{j2}, ..., w_{jk})$. PUNNs have several advantages, including increased information capacity and the ability to express strong interactions between input variables. Furthermore, it is possible to obtain upper bounds of the Vapnik-Chervonenkis (VC) dimension of PUNNs similar to those obtained for MLPs [19]. Despite these advantages, PUNNs have a major handicap: they have more local minima and more probability of becoming trapped in them [10]. The activation function of the PUNN considered in this work is given by:

$$f_{\text{PUNN}}(\mathbf{x}, \boldsymbol{\theta}) = \beta_0 + \sum_{j=1}^{m} \beta_j \prod_j(\mathbf{x}, \mathbf{w}_j)$$

with $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{w}_1, ..., \mathbf{w}_m)$, the softmax activation function being also considered:

$$g(\mathbf{x}, \boldsymbol{\theta}) = \frac{e^{f_{\text{PUNN}}(\mathbf{x}, \boldsymbol{\beta})}}{1 + e^{f_{\text{PUNN}}(\mathbf{x}, \boldsymbol{\beta})}}$$

In order to estimate the coefficients and the structure of the PUNNs that minimise the classification error function, an Evolutionary Algorithm (EA) has been selected. The algorithm is similar to the one proposed by Martínez-Estudillo et al. [16]. The population-based EA for architectural design and the estimation of real-coefficients have points in common with other EAs in the bibliography [3, 23]. The search begins with an initial population. This population is updated in each generation using a population-update algorithm, and is subject to the evolutionary operations of replication and mutation. Crossover is not used due to its potential disadvantages in evolving artificial networks [3]. For this reason, this EA belongs to the Evolutionary Programming (EP) paradigm. The general structure of the EA can be consulted in [16].

The adjustment of both the weights and structure of the PUNNs is performed by the complementary action of two mutation operators: parametric and structural mutation. Parametric mutation implies a modification in the coefficients (β_j) and the exponents (w_{ji}) of the model. Structural mutation modifies the topology of the neural nets, helping the algorithm to avoid local minima and increasing the diversity of the individuals trained. Five structural mutations are applied sequentially to each network [16]: neuron deletion, connection deletion, neuron addition, connection addition and neuron fusion.

4. Logistic Regression using Product Units (LRPU)

Logistic Regression using Product Units (LRPU) is a hybrid method that considers the EA presented in the previous section in order to obtain an EPUNN structure and hidden neuron weights that are accurate enough. When these are obtained, it applies the IRLS mechanism over the PU basis functions of the EPUNN selected. So the LRPU model composed only of PU basis function is given by:

$$f_{\text{LRPU}}(\mathbf{x}, \boldsymbol{\theta}) = \alpha_0 + \sum_{j=1}^{m} \alpha_j \prod_j(\mathbf{x}, \mathbf{w}_j)$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \mathbf{W})$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, ..., \alpha_m)$ and $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_m)$, with $\mathbf{w}_j = (w_{j1}, w_{j2}, ..., w_{jk})$. The coefficients $\mathbf{W}$ are given by the EA, and are not adjusted by the IRLS method. The IRLS method only optimises the linear part of the model, i.e., the $\boldsymbol{\alpha}$ coefficients.

5. Logistic Regression using Initial covariates and Product Units (LRIPU)

The LRIPU model used is a hybridisation of the LR model and the EPUNNs previously presented. The model extends LRPU, considering the initial covariates $\mathbf{x}$ of the problem. Its expression is given by:

$$f_{\text{LRIPU}}(\mathbf{x}, \boldsymbol{\theta}) = \alpha_0 + \sum_{j=1}^{m} \alpha_j \prod_j(\mathbf{x}, \mathbf{w}) + \sum_{j=1}^{k} \alpha_{(m+j)} x_j$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \mathbf{W})$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, ..., \alpha_m, \alpha_{m+1}, ..., \alpha_{m+k})$ and $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_m)$. The values adjusted with IRLS correspond to the $\boldsymbol{\alpha}$ vector, the coefficients $\mathbf{W}$ again being given by the EA.

6. Feature selection

In order to reduce the size of LRPU and LRIPU models, we propose two different feature selection methods. Both methods make use of the Wald statistic, which is a score function commonly considered in LR. The Wald test is a statistical test, used to check whether the effect of a covariate exists or not in the odd of an event. In other words, it tests whether an independent covariate has a statistically significant effect over the dependent variable. As a result, a critical value (p-value) is obtained for each variable, where the associated coefficient equal to zero is the null hypothesis (H_0) to be contrasted.

Backward Search Algorithm:
Apply IRLS over V, obtaining the $\boldsymbol{\alpha}$ coefficients and the associated CCR_{T}
$exit \leftarrow$ false
repeat
  for all v_i in V **do**
    Obtain Wald statistic of the variable v_i
    $p_i \leftarrow p$-value of the Wald test with $H_0 \equiv \alpha_i = 0$
  end for
  $v_{1\mathrm{st}} \leftarrow$ variable with maximum p_i
  $V^{'} \leftarrow V - v_{1\mathrm{st}}$
  Apply IRLS over $V^{'}$, obtaining the $\boldsymbol{\alpha}^{'}$ coefficients and the associated $CCR_{\mathrm{T}}^{'}$
  if $CCR_{\mathrm{T}} > CCR_{\mathrm{T}}^{'}$ **then**
    $v_{2\mathrm{nd}} \leftarrow$ variable with second maximum p_i
    $V^{'} \leftarrow V - v_{2\mathrm{nd}}$
    Apply IRLS over $V^{'}$, obtaining the $\boldsymbol{\alpha}^{'}$ coefficients and the associated $CCR_{\mathrm{T}}^{'}$
    if $CCR_{\mathrm{T}} > CCR_{\mathrm{T}}^{'}$ **then**
      $exit \leftarrow$ true
    else
      $V \leftarrow V^{'}$
    end if
  else
    $V \leftarrow V^{'}$
  end if
until $exit$=true

Figure 1. BS feature selection algorithm

6.1. Feature selection by a two-step Backward Search algorithm (BS)

The first method proposed starts with the full model with all the covariates, initial and PUs, pruning variables to the model sequentially and successively, until no further pruning can be made to improve the fit. At each step, the least significant covariate is selected in the discriminant function. The selected covariate is deleted if this does not reduce the fit. If it does, the second least significant covariate is considered. The procedure ends when none of the two chosen covariates is deleted. The pseudo-code associated with this algorithm is presented in Figure 1, where V is the current set of covariates and CCR_{T} is the Correct Classification Rate in the training set.

6.2. Feature selection by a two-step Simulated Annealing algorithm (SA)

The second method is based on the standard SA heuristic [12]. The algorithm is very similar to that presented in the previous subsection but, when the elimination of a variable results in a lower training CCR (CCR_{T}), it is accepted with a probability $e^{dif/T}$, where dif is the CCR_{T} difference between the model obtained using the variable and not using it, $dif = (CCR_{\mathrm{T}}^{'} - CCR_{\mathrm{T}})$, and T is the current temperature. The initial value for the temperature is $T = 0.01N$, where N is the number of covariates and PUs of the model. In each iteration, the temperature is updated with a $r = 0.2$ freezing factor. The pseudo-code associated with this algorithm is presented in Figure 2, where $U(0,1)$ is a random uniform variable in the interval $[0,1]$.

Simulated Annealing Algorithm:
Apply IRLS over V, obtaining the $\boldsymbol{\alpha}$ coefficients and the associated CCR_{T}
$exit \leftarrow$ false; $T \leftarrow 0.01\cdot$ Number variables
repeat
  for all v_i in V **do**
    Obtain Wald statistic of the variable v_i
    $p_i \leftarrow p$-value of the Wald test with $H_0 \equiv \alpha_i = 0$
  end for
  $v_{1\mathrm{st}} \leftarrow$ variable with maximum p_i
  $V^{'} \leftarrow V - v_{1\mathrm{st}}$
  Apply IRLS over $V^{'}$, obtaining the $\boldsymbol{\alpha}^{'}$ coefficients and the associated $CCR_{\mathrm{T}}^{'}$
  $dif \leftarrow (CCR_{\mathrm{T}}^{'} - CCR_{\mathrm{T}})$
  if $dif < 0$ **and** $U(0,1) > e^{dif/T}$ **then**
    $v_{2\mathrm{nd}} \leftarrow$ variable with second maximum p_i
    $V^{'} \leftarrow V - v_{2\mathrm{nd}}$
    Apply IRLS over $V^{'}$, obtaining the $\boldsymbol{\alpha}^{'}$ coefficients and the associated $CCR_{\mathrm{T}}^{'}$
    $dif \leftarrow (CCR_{\mathrm{T}}^{'} - CCR_{\mathrm{T}})$
    if $dif < 0$ **and** $U(0,1) > e^{dif/T}$ **then**
      $exit \leftarrow$ true
    else
      $V \leftarrow V^{'}$
    end if
  else
    $V \leftarrow V^{'}$
  end if
  $T \leftarrow 0.2T$
until $exit$=true

Figure 2. SA feature selection algorithm

7. Experiments

We have tested the described methodology in a real agronomical problem of precision farming, consisting of mapping weed patches in crop fields, through remote sensed data.

The study was conducted on a 42 ha sunflower field located in Andalusia, southern Spain, named Matabueyes (co-

ordinates 37.8ºN, 4.8ºW, WGS84). The field was naturally infested by *R. segetum*. Conventional-colour (CC, 400-700 nm) and colour-near infrared (CIR, 500-900 nm) aerial photographs of the field were taken in mid-May. Input variables included the digital values of all bands in each available image, that is: CC images responded to blue (B, 400-500 nm), green (G, 500-600 nm), and red (R, 600-700 nm) broad bands of the electromagnetic spectrum, and CIR images to G, R and near-infrared (NIR, 700-900 nm) bands. Further information about acquisition of aerial photographs, digitisation and orto-rectification was described in [18].

To train and validate the classification models, a random ground sampling procedure was carried out in mid-May when the aerial image was taken ensuring that all parts of the field area had an equal chance of being sampled with no operator bias [17]. We obtained 2,400 pixels as ground-truth pixels and georeferenced a total of 1,600 pixels in each phenological stage, where 800 pixels corresponded to R. segetum class, 400 pixels corresponded to the bare soil class and 400 corresponded to that of sunflower. The objective is the differentiation between *R. segetum* and weed-free (bare soil and sunflower) pixels.

The experimental design was conducted using a stratified 10-fold cross-validation procedure. The models compared in the different experiments are the following: firstly, the application of the IRLS algorithm over only the PU basis functions extracted from EPUNN model of the EP algorithm (LRPU) and over the same basic functions together with initial covariates (LRIPU). Secondly, the two different feature selection algorithms (BS and SA) are applied over both LRPU and LRIPU models (LRPU-BS, LRPU-SA, LRIPU-BS and LRIPU-SA). Moreover, all these models are compared to six recent machine learning classifiers: LR with attribute selection (SimpleLogistic), LR with a full logistic model (MultiLogistic), Logistic Model Trees (LMT), the C4.5 classification tree inducer, the naive Bayes tree learning algorithm (NBTree) and the AdaBoost.M1 algorithm with 100 maximum number of iterations (AdaBoost100) and using C4.5 as the base classifier. The description of these algorithms can be found in [14]. The EP algorithm was implemented using the Evolutionary Computation framework JCLEC [21] (`http://jclec.sourceforge.net`) and is available in the non-commercial JAVA tool named KEEL [2] (`http://www.keel.es`). The other algorithms are available as part of the WEKA machine learning workbench [22] and they have been applied using their default parameter values.

Performance of each model has been evaluated using the CCR in the generalisation set (CCR_{G}). In Table 1 we show the mean and the standard deviation of this CCR_{G} for a total of 100 executions and the mean and the standard deviation of the number of coefficients of the corresponding model. From the analysis of the LRPU model results, it can be concluded that the feature selection methods considerably reduce the number of coefficients, this difference being higher for the LRPU-SA method. The generalisation accuracy of the LRPU models is higher when features are selected with the BS method (LRPU-BS) and it is very similar to that obtained when the LRPU-SA method is used. A very similar behaviour is observed with respect to the LRIPU model: the accuracy is enhanced when using feature selection (especially when using the LRIPU-BS method) and the number of coefficients is significantly reduced (especially when using the LRIPU-SA method). When analysing both LRPU and LRIPU models and their feature selection variants, better results are obtained by the LRIPU-BS method and a similar accuracy is obtained by the LRIPU-SA methodology but with a lower number of connections.

Table 1. Statistical results (Mean and Standard Deviation, SD) of the $\mathbf{CCR_G}$ and the number of coefficients (#Coef.) obtained using the different methods proposed

	CCR_{G}	#Coef.
Method	Mean $\pm$ SD	Mean $\pm$ SD
LRPU	69.53 ± 3.58	19.66 ± 2.19
LRPU-BS	70.07 ± 3.47	17.21 ± 3.14
LRPU-SA	69.49 ± 3.45	13.69 ± 2.04
LRIPU	69.84 ± 3.57	23.08 ± 2.46
LRIPU-BS	70.32 ± 3.58	20.20 ± 3.45
LRIPU-SA	70.10 ± 3.58	15.56 ± 2.91

In Table 2 we have included the most accurate results from the models proposed, together with the results obtained by using the different WEKA algorithms considering 10-fold cross-validation. The number of coefficients used in each algorithm model has also been obtained and included in Table 2. The accuracy of the models proposed in this paper is higher than that obtained by SimpleLogistic, MultiLogistic and NBTree and very similar to that obtained by the rest of algorithms. Moreover, the number of coefficients of LRIPU-BS and LRIPU-SA is significantly lower than that of those models that obtain a similar accuracy (LMT, C4.5 and AdaBoost100), resulting in more interpretable models.

8. Conclusions

The feature selection methods (BS and SA) presented in this paper have demonstrated an important coefficient reduction for both the hybrid neuro-logistic models proposed in [8] (LRPU and LRIPU), yielding to a better or similar accuracy. In this way, more interpretable models have been obtained that can lead to a better understanding of the clas-

Table 2. Results obtained with the proposed methodologies compared to other machine learning algorithms

	CCR_{G}	#Links
Method	Mean $\pm$ SD	Mean $\pm$ SD
SimpleLogistic	65.86 ± 3.30	4.46 ± 0.50
MultiLogistic	66.14 ± 3.27	5.00
LMT	71.68 ± 3.61	181.04 ± 91.23
C4.5	70.74 ± 3.54	33.63 ± 8.39
NBTree	66.78 ± 4.01	20.72 ± 9.08
AdaBoost100	70.89 ± 3.56	108.42 ± 29.96
LRIPU-BS	70.32 ± 3.58	20.2 ± 3.45
LRIPU-SA	70.10 ± 3.58	15.56 ± 2.91

sification problem tackled. Moreover, the comparison of these models to six different very refined machine learning methods has established the LRIPU-BS and LRIPU-SA methods as very competitive models with a lower number of coefficients and has demonstrated their capability to analyse multispectral imagery for predicting *R. segetum* presence probability in the field of study, providing a useful tool for early Site-Specific Weed Management.

Acknowledgements

This work has been partially financed by TIN2005-08386-C05-02 and AGL-2005-06180-CO3-02, projects of the Spanish Inter-Ministerial Commission of Science and Technology and FEDER funds. The research done by P.A. Gutiérrez and J.C. Fernández has been financed respectively by the FPU (grant reference AP-2006-01746) and the FPI (grant reference BES-2006-12543) Predoctoral Programs (Spanish Ministry of Education and Science).

References

[1] MAPA. Spanish Ministry of Agriculture, Fisheries and Food, 2007.

[2] J. Alcala-Fdez, L. Sánchez, S. García, M. J. del Jesús, S. Ventura, J. M. Garrell, J. Otero, C. Romero, J. Bacardit, V. M. Rivas, J. C. Fernández, and F. Herrera. KEEL: A software tool to assess evolutionary algorithms for data mining problems. *Soft Computing*, 2008. , In Press.

[3] P. J. Angeline, G. M. Sauders, and J. B. Pollack. An evolutionary algorithm that constructs recurren neural networks. *IEEE Transactions on Neural Networks*, 5:54–65, 1994.

[4] R. B. Brown and S. D. Noble. Site-specific weed management: sensing requirements - what do we need to see? *Weed Science*, 53(2):252–258, 2005.

[5] P. Carranza-Cañadas, M. Saavedra, and L. García-Torres. Competition of ridolfia segetum and sunflower. *Weed Research*, 35(5):369–375, 1995.

[6] R. Durbin and D. Rumelhart. Products units: A computationally powerful and biologically plausible extension to backpropagation networks. *Neural Computation*, 1(1):133–142, 1989.

[7] P. A. Gutiérrez, F. López-Granados, J. M. Peña-Barragán, M. Jurado-Expósito, M. T. Gómez-Casero, and C. Hervás. Mapping sunflower yield as affected by *Ridolfia segetum* patches and elevation by applying evolutionary product unit neural networks to remote sensed data. *Computers and Electronics in Agriculture*, 60(2):122–132, 2008.

[8] C. Hervas-Martinez and F. Martinez-Estudillo. Logistic regression using covariates obtained by product-unit neural network models. *Pattern Recognition*, 40(1):52–64, 2007.

[9] D. W. Hosmer and S. Lemeshow. *Applied logistic regression (Wiley Series in probability and statistics)*. Wiley Interscience, New York; Chichester, September 2000.

[10] A. Ismail and A. P. Engelbrecht. Pruning product unit neural networks. In *Proceedings of the 2002 International Joint Conference on Neural Networks*, pages 257–262, 2002.

[11] M. Jurado-Expósito, F. López-Granados, J. L. González-Andújar, and L. García-Torres. Characterizing population growth rate of convolvulus arvensis in no-tillage systems. *Crop Science*, 45(21):2106–2112, 2005.

[12] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[13] P. Komarek and A. W. Moore. Making logistic regression a core data mining tool with tr-irls. In *Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM'05)*, pages 685–688, Washington, DC, USA, 2005. IEEE Computer Society.

[14] N. Landwehr and M. Hall. Logistic model trees. *Machine Learning*, 59:161–205, 2005.

[15] R. L. Major and C. T. Ragsdale. Aggregating expert predictions in a networked environment. *Computers & Operations Research*, 28(12):1231–1244, 2001.

[16] F. J. Martínez-Estudillo, C. Hervás-Martínez, P. A. Gutiérrez, and A. C. Martínez-Estudillo. Evolutionary product-unit neural networks classifiers. *Neurocomputing (In Press)*, 2008.

[17] R. M. McCoy. *Field Methods in Remote Sensing*. The Guilford Press, 2004.

[18] J. M. Peña-Barragán, F. López-Granados, M. Jurado-Expósito, and L. García-Torres. Mapping *Ridolfia segetum* moris patches in sunflower (Helianthus annuus L.) crop using remote sensing. *Weed Research*, 47(2):164–172, 2007.

[19] M. Schmitt. On the complexity of computing and learning with multiplicative neural networks. *Neural Computation*, 14:241–301, 2002.

[20] A. Srinivasan. *Handbook of Precision Agriculture*. Haworth Press Inc, 2006.

[21] S. Ventura, C. Romero, A. Zafra, J. A. Delgado, and C. Hervas. JCLEC: a Java framework for evolutionary computation. *Soft Computing*, 12(4):381–392, 2008.

[22] I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, second edition, June 2005.

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