ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 459-464

Neuro-Fuzzy Methodologies for the Clustering and the Reliability Estimation of Olive Fruit Fly Infestation

Bellei E.(1), Petacchi R.(2), Reyneri L.(1)

(1) Department of Electronics, Politecnico of Torino
C.so Duca degli Abruzzi 24. I-10129 Torino – ITALY
Phone (++39)011 564 4045 (++39)011 564-4038; fax (++39) 011 564 4099
e-mail : bellei@charlie.polito.it , reyneri@polito.it
(2) Agricultural Sector, Scuola Superiore S.Anna of Pisa
Via Carducci 40. I-56100 Pisa – ITALY
Phone (++39) 050 883442 ; fax (++39) 050 883215
e-mail : ruggero@sssup.it

Abstract. The present article describes the last results obtained from the application of neuro-fuzzy techniques in the study of Bactrocera Oleae infestation in the Liguria region olive grows [3]. The project "Applications of Neuro-Fuzzy Techniques in Agriculture" started on March 2000 with the monitoring and collection of data from a large number of oil farms. The main aim of the project was realized an area-wide Bactrocera Oleae monitoring network in order to administer IPM and to provide technical assistance in treatments to each farm. The way to reach this aim had been the creation of neuro-fuzzy systems for the extraction of infestation's features to make an appropriate classification with good labels to suggest treatments for each monitored farm, based also on the estimation of olive fly measure's reliability.
During the project, it has been proved that standard approaches to made forecast analyses on data referred to the growth of olive fly, give results less good and flexible than those obtained with the new analysis techniques like neuro-fuzzy methodologies, which are more adapted for non-linear and complex problems like the agronomic ones.

1. INTRODUCTION

The Bactrocera Oleae [5] is the key pest in olive grove agronomic system in Liguria; the modern strategies against this pest are focused on the reduction of insecticides thanks the application of a supervised control and IPM (Integrated Pest Management) strategies. The quality of an agrarian product (olive oil quality) deals with its geographic and physical characteristics but, for a good classification of the oil quality, many other factors shall be considered like growing and transformation. The olive fly infestation and the used defence techniques influence strongly the olive production and the oil quality; therefore, it is necessary to assess the infestation dynamics during the olive fruit fly growth.

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 459-464

The area wide monitoring networks has been established in order to supply technical assistance on olive fruit fly control. In monitoring points located in someone of the olive grove cultivated area, the technicians which follow the infestation dynamic, sample olive fruit, analyse infestation in laboratory, transmit information to a Centre of Data Elaboration and advise the farmers about the treatments that shall be provided.
The data collection was used to train and test neuro-fuzzy systems for the extraction of information and features of infestation. In fact, the ecological phenomena can be studied by using cognitive methods as neuro-fuzzy techniques for data characterization, classification and evaluation; in order to analyse large data set of infestation, a good approach is to realise an automatic descriptor of infestation data.
The infestation has been clustered in five different infestation indexes, which are expressed in per cent of total olive sampled: Active (eggs, I and II stage alive larvae), Dead (I and II stage dead larvae), Damaging (III stages larvae and pupae), Total (the sum of the above) and Target (eggs, I and II stage larvae).
The olive fruit fly dynamics in the considered fields has been monitored weekly, from July to November, by a single technician. The ecological system of olive fruit fly is very difficult to be represented by traditional methods that make use of mathematical equations due to the non-linearity and large variability of the phenomena; the present work has used the WRBF-n neural network [1] as an alternative method to the traditional ones.

1.2. CHARACTERIZATION of INFESTATION with WRBF-NETWORK

To make a classification of infestation it is important to describe, in a simple way, a curve using known math functions that allow extracting a limited numbers of parameters which can describe very well the curve. The characterization of curve [3,4] with the interpolation of 3 Gaussian curves (one for generation), has been made building a two layers feed-forward neural WRBF-n network with a training back-propagation algorithm. The network extracts for each curve 9 parameters: 3 centres (samples corresponding to the maximum infestation of generations), 3 heights of Gaussian (maximum levels of infestation) and 3 dispersions namely the width of curve (the whole period of each generation).
The network makes the characterization considering initially each measure with the same reliability (degree=0.9) and then associating to each measure the relative reliability computed applying suitable entomological rules, see par. 2. The results obtained are very encouraging in fact the standard error between real infestation and the output of the trained network is quite low. Here below the standard errors obtained by 3 farms in Liguria, without the reliability of measures.

Farm	Std error with WRBF-2	Std error with WRBF-3
F1	7.47 %	7.29 %
F2	2.34 %	2.1 %
F3	3.3 %	3.91 %

Table 1. The standard errors obtained with the characterization from [5].

2. NEURO-FUZZY SYSTEM to ESTIMATE RELIABILITY of OLIVE FRUIT FLY

The most used technique for data sampling consists in a random sample of one olive per plant on at least one hundred plants.
In field condition, the infestation monitoring was performed by a high number of technicians and could be often affected by further errors (high spatial heterogeneity, inexperience to detect mortality in younger larvae,..) therefore it is mandatory to established a method to evaluate the error and to associate a reliability degree to each measure.
Our interest is the detection of errors and the evaluation of data reliability that is requested to improve correctness of technical advice and to obtain a large set of good quality data to perform data analysis. The knowledge on Bactrocera Oleae development based on scientific studies and expert's experience on analysis of data quality, allow formulating heuristic rule to detect data errors.
The "point by point" comparison rules are based on the analysis of fruit fly infestation data between two subsequent monitoring at one-week distance; a way to understand the reliability level of an infestation measure is to identify two kinds of errors based on comparison of fruit fly infestation data among monitoring at one week distance.
The two kinds of errors are the unexpected increase and unexpected decrease of the infestation.
We implemented a neuro-fuzzy system, based on these rules, to automatically detect these errors; we validated the model comparing its performance with data analysis conducted by technicians. We used a fuzzy system to assess data reliability on the whole data set discussing the frequency of different kind of error, and suggesting how to improve monitoring procedures.
The agronimists created seven equations to detect the error; for each equation, a fuzzy system has been created by means of a three-layered Weighted Radial Basis Function (WRBF) neural network [2]. In this network the first hidden layer contains so many neurons as the number of rules (four), the second one is a normalization layer for the first step of de-fuzzyfication and the output layer has one linear neuron corresponding to the value of reliability.
Follows an example of one of the seven fuzzy system (one equation, four rules):
IF $(EH_{(t)} - (EH + P_l)_{(t-1)}$ is 'No Error' THEN $Data_{(t)}$Relaibility1 is "High"
IF $(EH_{(t)} - (EH + P_l)_{(t-1)}$ is 'Low Error' THEN $Data_{(t)}$Relaibility1 is "Medium"
IF $(EH_{(t)} - (EH + P_l)_{(t-1)}$ is 'Medium Error' THEN $Data_{(t)}$Relaibility1 is "Low"
IF $(EH_{(t)} - (EH + P_l)_{(t-1)}$ is 'Big Error' THEN $Data_{(t)}$Relaibility1 is "Very low"
where EH are exit holes, P_l are Live Pupae.
So doing, each measure is associated to seven reliability values (one for each equation) and the minimum of the obtained reliability values represents the final measure reliability (Figure 1).
The definitions of fuzzy values 'no error', 'low', 'medium' and 'high error' and 'very low', 'low', 'medium' and 'good' reliability has been defined considering the results of field experiment.

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 459-464

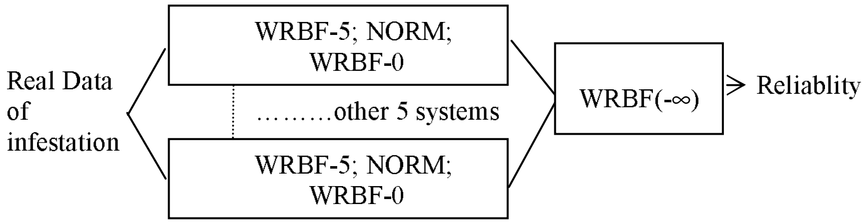

Fig. 1. Neuro-fuzzy system for reliability of measures

We have computed the fuzzy system on about 1600 observations on olive infestation made by technicians in Liguria Region during year 1999 and the results of reliability evaluation system have been compared with the crisp applications of the tests on the same data infestation performed by the agronomists.
In fact, three experts examined the data set obtained from monitored farms by assessing weekly data reliability and then we make a comparison among the different technician judgement and the fuzzy system result. The agreement among the results is very good and encourage for the future applications in other field.

	Human Expert Evaluation	
NF System Evaluation	Not reliable data	Reliable Data
Not reliable Data	91.67 %	8.33 %
Reliable Data	6.6 %	93.4 %

Table 2. Comparison among Experts and Fuzzy Systems.

There is a good advantage in using this automatic system because the time that the agronomist should spent to estimate the reliability of all measures for each farm is very large.
The estimated reliability is associated with each measure data and then used as a product factor in the training algorithm of the WRBF neural network for characterization; this results in a better characterization of infestation curve.
The standard error between the real infestation and the characterized curve from the trained network is generally worse than that obtained by using the same reliability for each measure, but the parameters which describe the curve are more reliable than the previous.

3. CLUSTERING OF INFESTATIONS

Before forecasting the behaviour of infestation and suggesting appropriate treatments, we proceeded with a classification of the cycles in well-defined and homogeneous groups, so to pick out groups of farms with similar behaviour and to study the elements of variability.
It is possible to see that, in standard climate conditions, the annual infestation of Bactrocera Oleae is characterized by the centres and by the level of each generation, but the micro-climate, the geographic position of farm and the adopted culture techniques are the elements which create a strong variability in infestation's

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 459-464

development. From a careful study of infestation development in different farms in Liguria, we evaluate the possibility to create a classification based on a number of clusters from a minimum of 7 to a maximum of 10.

For clustering, a WRBF neural network has been created with a variable number of inputs and one output that indicates what cluster the input data is part of. The WRBF neural network that was created for the clustering has two hidden layers of neuron: the first one contains as many neuron as the number of clusters, each one with Gaussian activation, while the second layer is a normalization layer that gives the target for not supervised back-propagation training [1] (Figure 2).

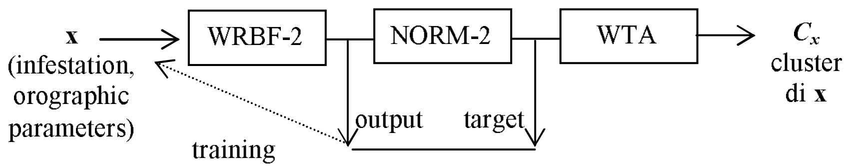

Fig. 2. Neuro-fuzzy system for clustering of Infestation.

Training network will give the infestation class for each farm and for each year, using a WTA neuron (NORM-∞). The input vector is composed of parameters obtained by the characterization with or without the meteo-geographic variables (like characterized median temperature, distance from sea, sun exposition, quote, longitude, latitude…).

The matrix of neural centres of hidden layer corresponds to the coordinates in the variables space of 'centroid' that characterized the group; so the training is like a "walk" of the centres of each cloud of points. We trained and tested the network with an iterative process which initializing the centers to random values and the weights to a constant value because we assumed all the groups with the same fixed dispersion.

Validation of this neural network required the research for similar characteristics among infestations of same class to optimize the classification. The studies made on the groups obtained with this network, showed that each one has farms with similar geographic features; this fact proof the efficiency of methodology chosen for this kind of problems. (Figure 3.).

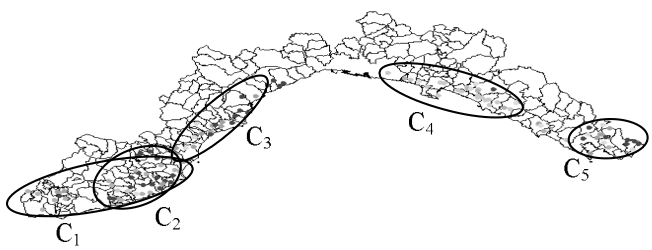

Fig. 3. Clustering results : farms in the same class (C_i) have similar geographic position

4. CONCLUSION

The availability of a relevant good quality data-set could be used for the extraction of ecological knowledge which is useful to define new fields of investigation.
The good results obtained by this study encourage the application of automatic procedure to assess data reliability, characterization and classification in other crops and fields of application.
Actually it is under study the spatial data heterogeneity to formulate space-temporal analysis of infestation for theoretical and practical purpose.
In this last period the WEB-based environment is developing for acquisition, processing, classification, forecast and management of disinfestations treatments.
The program is nearly finished, in fact the Web environment is available (some parts are under testing) to compute all neuro-fuzzy systems described on the olive fruit fly infestation, but it is adaptable to other agronomic system.

5. REFERENCES

[1] Reyneri L.: *Unification of Neural and Wavelet Networks and Fuzzy system*, IEEE Transactions on Neural Networks, Vol.10, No.4, July 1999
[2] Rajesh N. Davè, Raghu Krishnapuram.: *Robust Clustering Methods: A Unified View*, IEEE Transactions on Fuzzy Systems, Vol. 5, No.2, May 1997
[3] Bellei E., Guidotti D., Petacchi R., Reyneri L.M., Rizzi I.: *Applications of Neuro-Fuzzy Classification, Evaluation and Forecasting Techniques in Agricolture.* ESANN 2001, Bruges, Belgium, April, 2001.
[4] Bellei E., Guidotti D., Petacchi R., Reyneri L.M., Rizzi I.: *Evaluation, Classification and Clustering with Neuro-Fuzzy Techniques in Integrated Pest Management.* IWANN 2001, Granada, Spain, June, 2001.

Acknowledgement

This work has been conducted within a co-financed Italian MURST 1999 project "Applications of forecasting techniques in Agriculture".
We are grateful to the Regional Service of Applied Agro-meteorology of Liguria for providing data and for the fruitful collaboration within the Regional Project of Olive Oil Quality Improvement co-founded by European Union.