Neurocomputing 000 (2017) 1–12

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Treating stochasticity of olive-fruit fly's outbreaks via machine learning algorithms☆

Romanos Kalamatianos[a,*], Katia Kermanidis[a], Ioannis Karydis[a,b], Markos Avlonitis[a]

[a] Department of Informatics, Ionian University, Corfu 49132, Greece
[b] Creative Web Applications P.C., Corfu, 49131, Greece

ARTICLE INFO

Article history:
Received 17 January 2017
Revised 5 July 2017
Accepted 31 July 2017
Available online xxx

Keywords:
Olive fruit fly
Machine learning
Population prediction
Classification
Naïve Bayes
Nearest neighbors
Decision trees
Random forests
Support vector machines
Neural networks

ABSTRACT

Olive fruit fly trap measurements are used as one of the indicators for olive grove infestation, and therefore, as a consultation tool on spraying parameters. In this paper, machine learning techniques are used to predict the next olive fruit fly trap measurement, given input knowledge of previous trap measurements as well as an attribute that acts as a correlation model between the temperature and the development of a pest's population, known as the Degree Day model. This is the first time the Degree Day model is utilized as input in classification algorithms for the prediction of olive fruit fly trap measurements. Various classification algorithms are employed and applied to different environmental settings, in extensive comparative experiments, in order to detect the impact of the latter on olive fruit fly population prediction.

1. Introduction

The olive fruit fly is a pest that has been recorded to infest solely the olive fruit since at least the third century BC [2]. Such infestation causes great damage to the production of both olive oil and table olives [3] in many olive oil producing countries, including Greece. The olive fruit fly is active during the summer and reaches succesive population peaks during autumn, while during the winter and in the first months of spring it hibernates, until environmental conditions are favorable for it to re-emerge [2].

The population growth of the olive fruit fly and, by extension, the level of infestation of an olive grove are affected by various environmental factors. A short non-exhausting list of these factors can be summarized to: temperature, humidity, fruit bearing, olive grove orientation, olive grove variety, spatial diffusion, interaction between neighboring micro-climates [4–6]. The net result of all these factors is to introduce an spatiotemporal stochasticity in the evolution of olive fruit fly population. While an earlier model proposed by Avlonitis et al. [7] indeed addressed some of the aforementioned complex factors, such as spatial evolution of olive fruit fly, the robust modeling of spatiotemporal evolution can only be achieved by means of the stochastic generalization of the well known logistic equation for the olive fruit fly population, as shown in

$$\frac{\partial p}{\partial t} = \beta p(1-p) + c\frac{\partial^2 p}{\partial x^2} + g(p)\delta p \tag{1}$$

where p is the population density, β is the rate of increase, $c\frac{\partial^2 \cdot p}{\partial x^2}$ is the diffusion term in space and $g(p)\cdot\delta p$ models the spatiotemporal stochasticity, $g(p)$ being the corresponding noise amplitude. Within this context, the induced randomness in the time and space of the olive fruit fly outbreak emerges as one of the most crucial product parameters[1].

Population control of the olive fruit fly can be achieved through spraying of the olive trees, either with localized bait or universally [2,9] at an olive grove. However, in order for the spraying to have effect, it has to be applied when conditions are appropriate. Two factors indicate when spraying should commence [9]: (a) the ripeness level of the olive fruit, as the fruit needs to be ripe in order for it to be susceptible to the olive fruit fly and (b)

☆ This work is an extension of [1].

* Corresponding author.
E-mail addresses: c14kala@ionio.gr (R. Kalamatianos), kerman@ionio.gr (K. Kermanidis), karydis@ionio.gr, jonjon@cwa.gr (I. Karydis), avlon@ionio.gr (M. Avlonitis).

[1] For a recent stochastic model predicting population outbreaks the interested reader is referred to [8].

https://doi.org/10.1016/j.neucom.2017.07.071
0925-2312/© 2017 Published by Elsevier B.V.

the population of the fly, i.e. when a certain population threshold (recorded via sampling) is exceeded. Sampling is achieved through McPhail traps or yellow sticky traps [2,9]. The infestation threshold is usually set to seven olive fruit flies per trap per week during the summer while decreased to five olive fruit flies per trap per week during autumn [9]. In both cases each trap covers an area of 77.000 m^2 or approximately 1.000 olive trees.

The aim of this paper is to predict future olive fruit fly trap measurements, and by extension olive fruit fly infestations/outbreaks, using machine learning algorithms.

1.1. Motivation & contribution

The effect of insects that produce harm to humans' concerns and especially on crops, are of great significance. Pests infect and feed from the fruits and grains of agricultural goods, thus greatly reducing their value. In turn this leads to loss of both alimentary raw material as well as invested funds [10,11].

Moreover, existing research on olive fruit fly has focused mainly on aspects such as the biology, ecology, management, and impact on olive production. On the other hand, infestation prediction has yet to receive significant attention, despite the wide availability of pests' and contextual data from olive groves as well as the positive effects such a prediction could bring on treating olive fruit fly infestations and thus ameliorating the olive fruit production.

To address these requirements, our previous work [1] presented extensive experimentation with various classification algorithms, on different environmental settings, in order to detect the impact of environmental parameters on olive fruit fly population prediction. Therein, the proposed feature vector consisted of environmental parameters, specifically temperature and information about previous trap measurements. Nevertheless, the feature vector utilized in [1] was somewhat generic: it did not take into account the specifics of the olive fruit fly's bio-cycle, but rather incorporated parameters that affect it. Moreover, while the experimental evaluation was indeed promising, the distribution of the target class in the data presented a high imbalance, leading to poor performance in the prediction of the minority class.

This work significantly extends [1] by

1. Proposing a new feature that closely addresses requirements of the olive fruit fly. We incorporated the Degree Day model as feature, which correlates the environmental temperature with development of an organism.
2. Conducting and presenting promising experimental results with the aforementioned new feature. We experimented on three feature sets, where each feature set varied in the number of attributes, having as a constant attribute the aforementioned feature. Finally, the algorithms utilized therein were applied on four different experimentation methods.
3. Employing various means to address the class imbalance problem. Specifically, we utilized synthetic oversampling to the dataset via the SMOTE (Synthetic Minority Oversampling Technique) technique. Furthermore, we experimented with meta-learners. Finally, the three-fold bin of the trap measurement related attributes was reduced to two by merging two bins that ultimately served the same purpose.

The rest of the paper is organised as follows: Section 2 presents background information and related work, while Section 3 discuses the methodology utilized for the collection of data from environmental sensors and olive fruit fly traps, the features selected and extracted from the raw data as well as the class imbalance problem identified. Next, Section 4 details with the experimental setup and the experimental results obtained. Finally, the paper is concluded in Section 5.

2. Background & related research

This Section details necessary background information on machine learning methods as well as related existing research on olive fruit fly infestation prediction.

2.1. Machine learning algorithms

A number of classification algorithms exist that are suitable for the purposes of experimentation on the theme of this work. As classification approaches to olive fruit fly infestation prediction are extremely limited in number in the literature, the choice of classification algorithms has to a large extent been based on exploratory criteria, with the aim to cover varying learner families, including meta-learning. The following machine learning algorithms were used:

- J48 [12], a decision tree induction algorithm and it is a version of C4.5, an earlier algorithm developed by J. Ross Quinlan [13].
- Sequential Minimal Optimization or SMO [14], an ameliorated algorithm 100 for training support vector machines.
- Naïve Bayes [12], a probabilistic classifier based on the assumption of conditional independence [15].
- Random Forest [16], a meta-learning classification algorithm that runs iteratively.
- AdaBoost [17], another meta-learning algorithm.
- Ibk [12], an implementation of the k-nearest neighbor algorithm.
- Multilayer Perceptron, an artificial neural network [18].

Very often machine learning algorithms face the problem of over-fitting, i.e. they cope optimally when evaluated on the training data, but poorly on new unseen test data. Usually, the more complicated the model generated by a learning algorithm, the more prone it is to over-fitting. In our experiments, we dealt with the problem of over-fitting by

- Applying post-pruning, i.e. sub-tree raising, to the trees generated by the tree-based classifiers (J48 and RandomForest),
- Employing a low-complexity polynomial function in the Support Vector Machine kernel, i.e. a first-degree polynomial,
- Choosing a low-complexity perceptron structure, that includes only one hidden layer, and investigating its performance with varying numbers of nodes, starting from very low-complexity of only one node, to 15 nodes,
- Investigating other neural network structures of low complexity with varying number of inputs and hidden layers, thus of varying complexity.

As far as the problem of outliers is concerned, among all the attributes in our proposed feature vector, only one attribute had numeric values. After analysis of the numeric attribute no instances were found with an unusual/strongly deviating value. Therefore our feature vector space included no significant cases of outliers' instances.

2.2. State-of-the-art research on fruit fly infestation prediction

Related research indicates numerous attempts to simulate the population dynamics of the olive fruit fly as well as the prediction of outbreaks. Comins & Fletcher [19] developed a simulation model which predicted the phenology and dynamics of the olive fruit fly using field data. Pommois et al. [20] and Bruno et al. [21] used a cellular automata model to simulate the spatiotemporal infestation of olive groves by the olive fruit fly. Gilioli & Pasquali [22] used an individual-based model to model the development of the olive fruit fly numerically. Avlonitis et al. [7] proposed an evolution equation based on the dispersion of the olive fruit fly to express population outbreaks. In [23], Gutierrez et al. developed a

Please cite this article as: R. Kalamatianos et al., Treating stochasticity of olive-fruit fly's outbreaks via machine learning algorithms, Neurocomputing (2017), https://doi.org/10.1016/j.neucom.2017.07.071