Computers and Electronics in Agriculture 136 (2017) 86–96

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

An improved moth flame optimization algorithm based on rough sets for tomato diseases detection

Aboul Ella Hassanien [c,a], Tarek Gaber [b,c,*], Usama Mokhtar [c,d], Hesham Hefny [d]

[a] Faculty of Computers and Information, Cairo University, Egypt
[b] Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt
[c] Scientific Research Group in Egypt (SRGE), Egypt[1]
[d] Inst. of Stat. Studies and Res., Cairo University, Egypt

ARTICLE INFO

Article history:
Received 23 September 2016
Received in revised form 16 January 2017
Accepted 24 February 2017
Available online 14 March 2017

Keywords:
Moth flame optimization
Rough set theory
Particle Swarm Optimization (PSO) and Genetic Algorithms (GA)
Tomato's disease

ABSTRACT

Plant diseases is one of the major bottlenecks in agricultural production that have bad effects on the economic of any country. Automatic detection of such disease could minimize these effects. Features selection is a usual pre-processing step used for automatic disease detection systems. It is an important process for detecting and eliminating noisy, irrelevant, and redundant data. Thus, it could lead to improve the detection performance. In this paper, an improved moth-flame approach to automatically detect tomato diseases was proposed. The moth-flame fitness function depends on the rough sets dependency degree and it takes into a consideration the number of selected features. The proposed algorithm used both of the power of exploration of the moth flame and the high performance of rough sets for the feature selection task to find the set of features maximizing the classification accuracy which was evaluated using the support vector machine (SVM). The performance of the MFORSFS algorithm was evaluated using many benchmark datasets taken from UCI machine learning data repository and then compared with feature selection approaches based on Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) with rough sets. The proposed algorithm was then used in a real-life problem, detecting tomato diseases (Powdery mildew and early blight) where a real dataset of tomato disease were manually built and a tomato disease detection approach was proposed and evaluated using this dataset. The experimental results showed that the proposed algorithm was efficient in terms of Recall, Precision, Accuracy and F-Score, as long as feature size reduction and execution time.

1. Introduction

Plants are very crucial source of food and energy for humankind. Plant diseases can cause major economical, and ecological losses as well as reduction in both quantity and quality of agricultural products. Therefore, diagnosing and detecting plant diseases in a timely an accurate way is very important. Usually, the observation of experts using their naked eyes is the traditional approach followed in practice for the diagnosing and detection of plant diseases. Moreover, in some developing countries, small farmers could find difficulties to get experts making consulting these experts very expensive and time consuming. This could lead to the spreading of the disease into all crops. Thus, automatic/computer-based plant diseased detection approaches are of high importance.

The automatic detection system usually consists of two main phases. Firstly, the plant leaf image is captured using a digital camera. Secondly, the detection and classification of leaf diseases can be achieved through different steps: extracting the infected region, computing some features representing each disease and they classify these features to identify the diseases. The importance of automatic diagnosing and detection of plant diseases emerges as it could support benefits in monitoring big fields of crops, hence provide automatic detection of diseases based on the symptoms which appear on the plant leaves (Mokhtar et al., 2015).

In last years, automatic detection of plant diseases attracts many researchers in different fields. Bauer et al. (2009) proposed an approach for the automatic classification of leaf (i.e.,sugar beet) diseases using high resolution multi-spectral and stereo images. In Weizheng et al. (2008) introduced a new fast and accurate

[1] http://www.egyptscience.net.

* Corresponding author at: Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt.

E-mail addresses: aboitcairo@gmail.com (A.E. Hassanien), tmgaber@gmail.com (T. Gaber), usamamokhtar@yahoo.com (U. Mokhtar), hehefny@hotmail.com (H. Hefny).

http://dx.doi.org/10.1016/j.compag.2017.02.026
0168-1699/© 2017 Elsevier B.V. All rights reserved.

approach for grading plant diseases using computer image processing technique. They first used Otsu method to extract the leaf region, and then used Sobel operator to detect edges of the diseased spot. Finally, plant diseases are graded through the information of the quotient of disease spot and leaf areas as indicator. In another study Naidu et al. (2009) suggested a method to identify virus infected grapevine using the discriminant analysis and they obtained a maximum accuracy of 81% of the classification results. Also, cotton diseases (Camargo and Smith, 2009) were automatically identified using preprocessing operation and the use of SVM classifier to identify visual symptoms of cotton diseases. Moreover, in Liu et al. (2013) a new method for wheat disease identification using image recognition was proposed. In this method, after computing features of diseased region of leaf images, samples are trained and recognized using the RBF-SVM classifier. In Phadikar et al. (2012) to classify the leaf brown spot and the leaf blast diseases of rice plant, an automated system has been developed. This system is based on the morphological changes of the plants caused by the diseases and used the Bayes and SVM classifiers in the disease identification. Also an approach to detect the symptoms of nutrient diseases (Asraf et al., 2012) was suggested and it is based on the vision system and pattern recognition.

The feature selection process is one of the most important tasks for pattern recognition and classification systems, e.g. plant disease detection system. The main goal of this process is to find a minimal feature subset from a problem domain such that to give a high accuracy in representing the original features (Dash and Liu, 1997). It improves the predictive accuracy of algorithms by reducing the number of features, removing irrelevant, noisy and redundant features. It is also helps in the improvement of the classification performance. The feature selection mechanism has been successfully employed to effectively solve classification problem in various areas, such as bioinformatics (Salama et al., 2016), image processing (Salama et al., 2011), data mining (Lutu and Engelbrecht, 2010), pattern recognition (Sweidan et al., 2015), medical diagnosis (Ali et al., 2014; Sayed et al., 2016).

Different techniques were used to achieve feature selection. This includes the rough set theory (Pawlak, 1982) and bio-inspired techniques. The basic idea of using rough set-based for feature selection is to generate all possible feature reductions and then choose the one with minimal cardinality (Komorowski and Øhrn, 1999). The rough set has already used to accomplish a features selection task in different area such as Emary et al. (2014), Zawbaa et al. (2015) and Aziz and Hassanien (2016). Also, many bio-inspired methods have been used for feature selection process and these include Genetic Algorithm (GA) (Lu et al., 2008; Oliveira et al., 2003), ant colony optimization (ACO) (Basiri et al., 2008; Aghdam and Ghasem-Aghaee, 2008), Bat Algorithm (BA) (Nakamura et al., 2012; Rodrigues et al., 2014) and Grey Wolf Optimizer (GWO) (Emary et al., 2015).

Efforts have been targeted to combine the RS approach with bio-inspired algorithms to improve the performance. Bello et al. (2005) proposed an feature selection approach which integrates Ant Colony System with rough set. The approach firstly generates a number of ants which are placed randomly on the graph and then they traverse edges probabilistically until a traversal stopping criterion is satisfied to output the best rough set reduct. This method achieved a high ratio in features reduction but the classification accuracy and execution time are not good enough. Similar to the Bello's approach (Bello et al., 2005), Wang et al. (2007) introduced an approach integrating between rough set and the Particle Swarm Optimization (PSO) to achieve the feature selection task. They followed the same idea but only applied PSO instead of ACS. Wang's approach was able to find the optimal reducts on most of the used datasets and minimizing the execution time. In another effort, Guo et al. (2010) proposed an approach combining between Genetic Algorithm, GA, and rough set for the feature selection. Firstly, rough set was used to carry out the feature selection, then to find the optimal subset in the remaining feature subset, they used the GA improved with Population Clustering. The SVM (Support Vector Machines) was then applied to evaluate the effectiveness of the selected feature subset.

In this paper we proposed a Moth-Flame Optimization (MFO) and rough set (MFORSFS) approach for automatically detecting some kinds of tomato disease. The tomato was chosen to be the application of the automatic disease detection in this study because of its importance. It is ranked number one among 40 vegetables/fruits in terms of "relative contribution to human nutrition" and contains a high nutrition value. To achieve tomato disease detection, feature selection is a important phase. Thus, we first have introduced a new feature selection technique based on MFO and Rough Set called MFORSFS. This MFORSFS was evaluated to prove its robustness and then we have used in the detection of the tomato diseases. The proposed MFORSFS algorithm was compared against using (1) Particle Swarm Optimization (PSO) and (2) Genetic Algorithm (GA) with the rough sets. The results showed that the MFORSFS gave a higher accuracy of classification results while preserve low number of features compared to the other two optimization algorithms. The MFORFS was then used to select the best features describing the tomato leave and this helped in achieving a high accuracy comparing to the most related work.

The rest of this paper is organized as follows: Section 2 gives an overview of the moth flame optimization and rough sets. Section 3 presents the details of the proposed system. In Section 4, experimental results and discussion are given. Finally in Section 5, conclusions and future work are presented.

2. Preliminaries

2.1. Gabor features

Gabor filter-base method is an effective method for extracting texture feature. It has been used in many applications such as biometrics and segmentation. Gabor filters are known as convolution kernel, the product of a cosine and Gaussian functions. It enjoys the characteristic of specified orientation and spatial frequency. The 2-D Gabor filter is like a local band-pass filter with some localization properties in the spatial and frequency domain. Gabor filter is proved his efficiency in characterizing texture features (Grigorescu et al., 2002), like in our case: extracting texture features from tomato's leaves.

A 2D Gabor function $g(x,y)$ is defined as follows:

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi jWx\right] \tag{1}$$

where σ_x and σ_y characterize the spatial extent and frequency bandwidth of the Gabor filter, and W represents the frequency of the filter. Let $g(x,y)$ be the mother generating function for the Gabor filter family. A set of different Gabor functions $g_{m,n}(x,y)$ can be generated by rotating and scaling $g(x,y)$ to form an almost complete and non-orthogonal basis set, that is,

$$g_{m,n}(x,y)) = a^{-2m} g(x',y')) \tag{2}$$

where $\acute{x} = a^{-m}(x\cos\theta_n + y\sin\theta_n), \acute{y} = a^{-m}(-x\sin\theta_n + y\cos\theta_n)$, $a > 1, \theta_n = n\pi/K, m = 0,1,\ldots,S-1$, and $n = 0,1,\ldots,K-1$. Parameter S is the total number of scales, and parameter K is the total number of orientations. So, S and K represents the total number of generated functions.

Given an image $I(x,y)$, its Gabor-filtered images are