2020 17th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)
Mexico City, Mexico. November 11-13, 2020

A Clustering Method Based on the Artificial Bee Colony Algorithm for Gas Sensing

J. Enríquez-Gaytán, F. Gómez-Castañeda, J.A. Moreno-Cadenas, L.M. Flores-Nava
Electrical Engineering Department, Cinvestav-IPN, Mexico City, Mexico
phone no. (52) 55 5747 3800 Ext. 6261
e-mail: {jenriquezg, fgomez, jmoreno, lmflores}@cinvestav.mx

Abstract— The authors present a method for finding the centers of clusters of complex data, where the classes are known a priori. This method uses the metaheuristic algorithm of the Artificial Bee Colony (ABC) algorithm, where the objective function is formulated according to the optimal K-means criteria. The used data come from a characterization laboratory, which deals with sensing gases using arrays of metal oxide devices. It is selected ethylene, methane and carbon monoxide for this clustering method. The main parts of the ABC algorithm are described in detail.

Keywords—Artificial Bee Colony algorithm; K-means; MOX gas sensor; gas sensing

I. Introduction

The usage of metal oxide (MOX) gas sensors has become an industrial standard, where the security is often an issue. This kind of sensors transduce the concentration of gases as a change of resistivity, which with the help of a reading circuit an analog value is provided for its numerical processing. The drawback in using MOX gas sensors comes from their intrinsic variability, which is generally taken into account with a calibration procedure [1]. Other techniques based on a set of MOX gas sensors can reduce the transduction error with simple computational approaches [2]. In this work, we find the center of clusters as the representative point of a group of MOS gas sensors under a wide range of concentrations of toxic and dangerous gases namely, ethylene, methane and carbon monoxide.

Swarm algorithms belong to the Metaheuristic algorithms methods. These algorithms are inspired by biological social systems leading to, for example, the Ant Colony Optimization (ACO) Algorithm [3], the Particle Swarm Optimization (PSO) Algorithm [4], and the Artificial Bee Colony (ABC) Algorithm. In recent years, they have been used to solve complex optimization problems. In this paper, we use the ABC algorithm for the localization of the centers of the clustering in the problem of detection of dangerous gases. Clustering is a generic algorithmic method applied in data analytics and whose aim is to find the members of the clusters and also their centers [3]. Therefore, this task is divided into two processes namely, one for finding the members of clusters and the other one for finding the center of the clusters. In this paper, the classes of the available data is already known so, the algorithm for determining their number is not necessary.

In particular, the core of the classic K-means clustering algorithm, which is related to the minimization of the aggregation of all the Euclidian distances between the data of each class and their centers, is performed in this work with the ABC algorithm; in it, the initial center of clusters are the sources of food, on which the onlooker and employed bees exploit. As the algorithm goes on the third kind bees appear doing exploration and reporting new and better sources of food. In the last epochs of the ABC algorithm, the present sources of food become the optimal center of the clusters. This strategy in the ABC algorithm reduces notably the computational complexity of this clustering task. The results are compared with the PSO and the K-means algorithms.

This paper is divided in sections. Section II deals with introducing the k-means clustering method. Section III explains the ABC algorithm. The objective function for the clustering task is formulated in section IV. The origin of the experimental data is exposed in Section V. Section VI depicts the clustering results. And, Section VII ends the paper with conclusions.

II. K-means Clustering

The K-means clustering algorithm finds the members of the classes into a data set and also their centers, where the number of classes are established as an initial parameter. Below are the main lines of this algorithm, where x_n constitute the data set.

K-means algorithm

1. Randomly initialize the clustering centers.

$$\boldsymbol{\mu_1, \mu_2, \dots, \mu_k} \in \mathbb{R}^2$$

Repeat until the value no longer changes.

2. For each $\boldsymbol{n^{th}}$ data point perform:

$$\boldsymbol{c_{nk}} = \begin{cases} \mathbf{1} & \boldsymbol{if\ k = arg\,min_j} \left|\left| \boldsymbol{x_n - \mu_j} \right|\right|^2 \\ \mathbf{0} & \boldsymbol{otherwise} \end{cases}$$

3. For each $\boldsymbol{k}$ do:

$$\boldsymbol{\mu_k} = \frac{\sum_n \boldsymbol{c_{nk} x_n}}{\sum_n \boldsymbol{c_{nk}}}$$

978-1-7281-8987-1/20/$31.00 ©2020 IEEE

Where the operator: $\arg min_j\, g()$ gives the value of j for which $g()$ attains it's minimum. The discrete variables, which appear as sub-fixes take care of counting the number of combinations of possible states. This algorithm suffers from being stacked into a "local minimum" so, it should be supported by some evaluation for goodness.

III. Artificial Bee Colony Algorithm

The ABC algorithm belongs to the metaheuristic algorithms category in the whole set of optimization methods available in engineering. It was proposed by Karaboga [5]. At present, it is also recognized as both computationally flexible and efficient. Its description follows below.

This algorithm has 4 phases and 3 types of bee, according to:

- Initialization
- Employed Bee
- On-looker Bee
- Explorer Bee

The ABC algorithm finds the set of values in the parameters of the objective function:$f(x)$, minimizing it. The variable x, which is in general a vector, defines the search space where the optimal solution might exist.

This algorithm is capable of finding the global minimum, which is associated to the most valuable source of food and also of abandoning from local minima.

Initialization. The bee colony has SN members and is divided into two groups. The first half refers to the employed bees and the other half to the on-looker bees. The food of sources i.e. the solution vectors are defined by (1), with dimension SN.

$$x = (x_1, x_2, \dots, x_{SN},) \in \mathbb{R}^{SN} \quad (1)$$

The search space is limited to the below interval:

$$l_i \leq x_i \leq u_i, \qquad i = 1, \dots, SN$$

Where: l_i and u_i are the lower and upper limits, respectively. The initialization starts with (2), which is in charge of making a homogenous distribution of the employed bees in the search space; it is considered as the food sources. Therefore, every employed bee will have its own food of source to be exploited.

$$x_{i,j} = u_{i,j} + \phi_{i,j}(u_{i,j} - l_{i,j}) \quad (2)$$

Where: $i = 1,2, \dots, SN,\ j = 1,2, \dots, D$, ϕ_{ij} is a random number in [0,1]. After this, the objective function is evaluated for all food sources.

Employed Bee Phase. This phase performs exploitation to the neighboring selected food source, using (3).

$$v_{i,j} = x_{i,j} + \theta_{i,j}(x_{i,j} - x_{k,j}) \quad (3)$$

Where: $k \in \{1,2, \dots, SN\}$ y $j \in \{1,2, \dots, D\}$, they are selected randomly and k is different from i. The parameter $\theta_{i,j}$ is a random number in [-1, 1]. After the evaluation of the objective function with the solution vector v_{ij}, a "greedy" process follows, which selects the best result against $f(x_{ij})$.

On-looker Bee phase. It uses the most abundant food source ever found as a pivot for exploring new food sources. This process is depends on the probability, which is associated to the food source (4).

$$p_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n} \quad (4)$$

In order to propose a new source of food, (5) is used.

$$v_{i,j} = x_{i,j} + \Phi_{i,j}(x_{i,j} - x_{k,j}) \quad (5)$$

Where: $k \in \{1,2, \dots, SN\}$ y $j \in \{1,2, \dots, D\}$; they are selected randomly and k is different from i. $\Phi_{i,j}$ is a random number in [-1, 1]. After the evaluation of the objective function with the solution vector v_{ij}, a "greedy" process follows, which selects the best result against $f(x_{ij})$.

Explorer Bee Phase. In this phase, the food sources whose performance has not improved after a number of trials L are substituted for a new food sources using (6).

$$x_{i,j} = u_{i,j} + \Phi_{i,j}(u_{i,j} - l_{i,j}) \quad (6)$$

Again, the search space is limited to the below interval:

$$l_i \leq x_i \leq u_i, \qquad i = 1, \dots, SN$$

Where: l_i and u_i are the lower and upper limits, respectively. Fig. 1 shows the flow diagram of the ABC algorithm, where the above phases take their place.

IV. Objective Function

This function has the purpose of evaluating if the center of the cluster has being found by the ABC algorithm. It happens when its graph "objective function value" versus "number of epoch" saturates at the lower level. The analytical expression for this function is formulated as the sum of all the Euclidean distances between the center of cluster and the individual samples taken from the data set. Below is the used objective function in this work.

$$\boldsymbol{f(\mu_j)} = \sum \left|\left|\boldsymbol{y_n} - \boldsymbol{\mu_j}\right|\right|^2 \quad (7)$$

Where: $\boldsymbol{y_n}$ and $\boldsymbol{\mu_j}$ are one sample of the analized data and the center of the cluster, respectively. In this case, the ABC algorithm takes $\boldsymbol{\mu_j}$ as the food of source in the space of search.

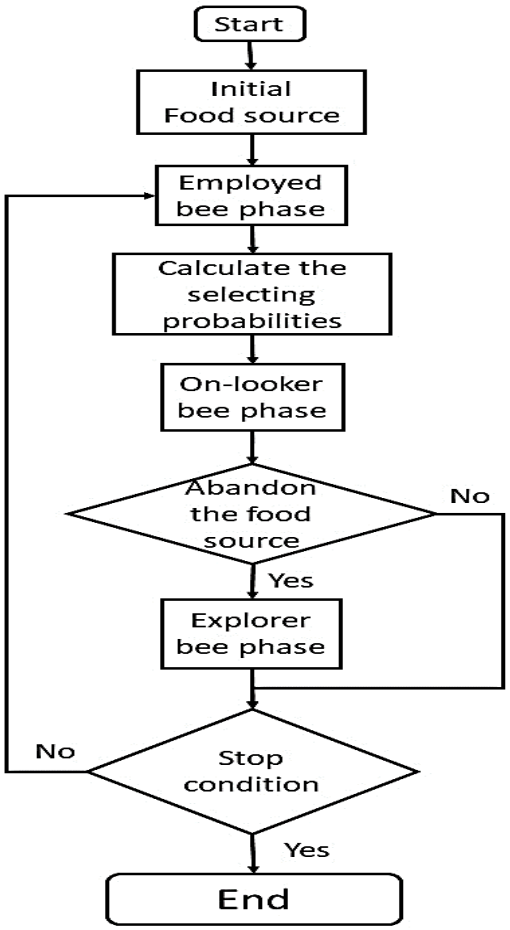

Fig 1. Flow diagram of the ABC algorithm.

V. Experimental Data

The experimental data, which serve as a means to demonstrate the present clustering method, are taken from an academic repository at UC, Santa Barbara, USA [6].

The sensing devices are of the TGS-series by Figaro Engineering Inc. [7]. They are based on metal-oxide semiconductor films and react by exposition to industrial-standard gases changing their electrical resistivity. The sensitivity is reset by a heater unit, which is integrated aside. The electrical response of the sensor is reported in data sheets as a mean value from sampling lots with approximate models. The user is responsible for dealing with their actual variability with adaptive electronics or software, for example. Another way to cope with this issue is via pattern recognition methods or intelligent algorithms. In particular, the clustering method in this work that combines the Principal Component Analysis technique (PCA) and the ABC algorithm contributes to this research area. From [8], the experimental data related to ethylene, methane, and carbon monoxide were clustered as a solution to their detection.

VI. Clustering Results

Fig. 2 depicts the computing processes for doing clustering with the proposed method.

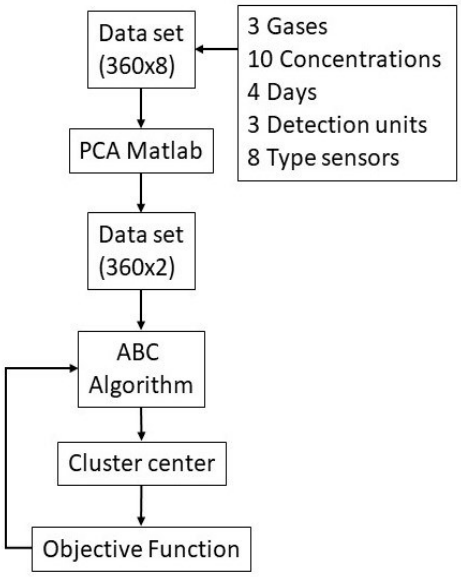

Fig. 2 Proposed clustering method.

The PCA stage reduces the original data dimension from 360x8 to 360x2 and allows visualizing the intrinsic clusters in it. Then, the ABC algorithm finds the center of the clusters.

The original data set comes from detecting 3 dangerous gases namely, ethylene, methane and carbon monoxide, whose concentration corresponds to 10 values inside a standard interval. It was also considered for these data those created in 4 days with one experimental session per day. The experimental set-up had 3 identical electronic boards, each one with an array of 8 different-part-number Figaro sensors.

Table 1 summarizes the distances of the clusters given by the ABC, PSO, and K-means algorithms, where the PSO and the ABC algorithm have remarkable results.

Table 1. Results comparison of optimization algorithms.

Algorithm	Total sum of distances
K-means	75.7363
ABC algorithm	62.7741
PSO algorithm	62.7747

Fig. 3 shows the graphical outcome of this clustering method, where the axes belong to the first two principal components: PC1 and PC2, computed with the PCA function in Matlab.

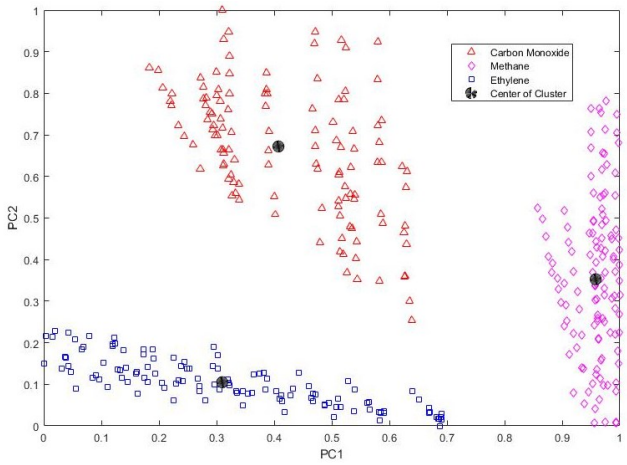

Fig. 3 Clustering of ethylene, methane and carbon monoxide.

For reason of completeness, Figs. 4-6 show the evolution of the used objective function by the ABC algorithm, when it computes the center of clusters marked in Fig. 3.

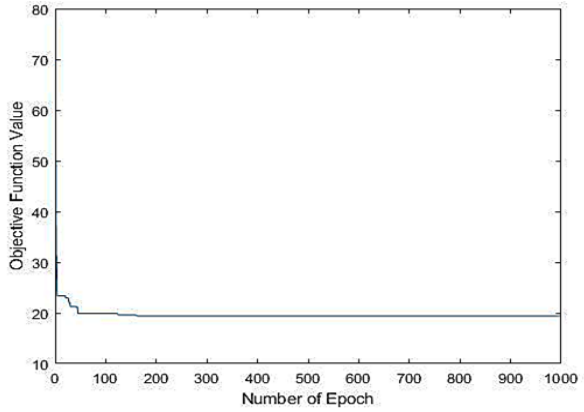

Fig. 4 Evolution of the objective function for ethylene.

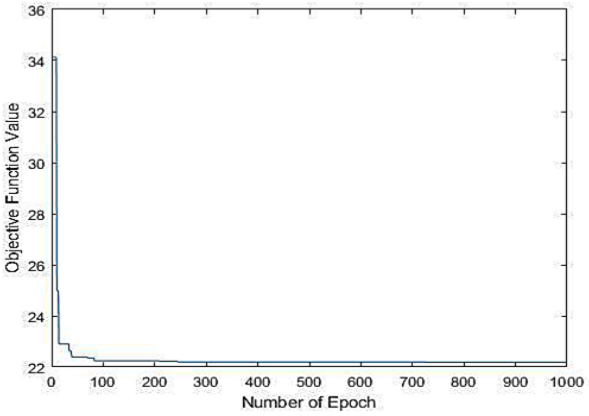

Fig. 5 Evolution of the objective function for methane.

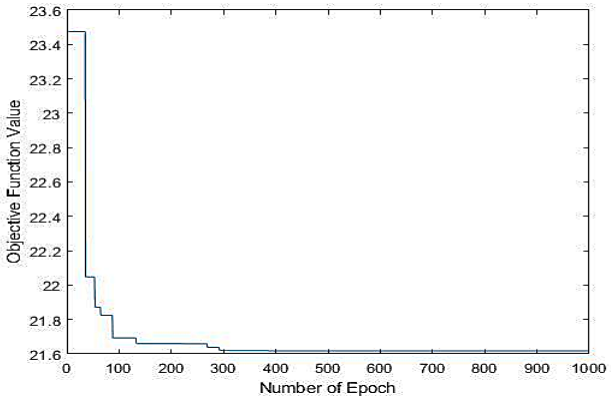

Fig. 6 Evolution of the objective function for carbon monoxide.

VII. Conclusions

The proposed combination of both the PCA method and the ABC algorithm as an alternative way for approaching the K-means clustering task, was demonstrated with a complex data set. The detection of 3 dangerous gases was solved efficiently, via vusualizing their clusters and respective centers. In this vehicle problem the high variability of actual MOX gas sensor divices makes the original data set difficult to analize.

References

[1] E. Martinelli, G. Magna, A. Vergara, and C. Di Natale, "Cooperative classifiers for reconfigurable sensor arrays," *Sensors Actuators, B Chem.*, vol. 199, pp. 83–92, 2014.

[2] J. E. Haugen, O. Tomic, and K. Kvaal, "A calibration method for handling the temporal drift of solid state gas-sensors," *Anal. Chim. Acta*, vol. 407, no. 1–2, pp. 23–39, 2000.

[3] W. Gao, "Improved Ant Colony Clustering Algorithm and Its Performance Study," *Comput. Intell. Neurosci.*, vol. 2016, p. 4835932, 2016.

[4] M. Zhao, H. Tang, J. Guo, and Y. Sun, "Data Clustering Using Particle Swarm Optimization BT - Future Information Technology," , pp. 607–612, 2014.

[5] D. Karaboga and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," *Appl. Soft Comput. J.*, vol. 8, no. 1, pp. 687–697, 2008.

[6] J. Fonollosa, L. Fernández, A. Gutiérrez-Gálvez, R. Huerta, and S. Marco, "Calibration transfer and drift counteraction in chemical sensor arrays using Direct Standardization," *Sensors Actuators B Chem.*, vol. 236, pp. 1044–1053, 2016.

[7] "FIGARO ENGINEERING INC." [Online]. Available: https://www.figaro.co.jp/en/. [Accessed: 09-Aug-2020].

[8] J. Enríquez-Gaytán, F. Gómez-Castañeda, L. M. Flores-Nava, and J. A. Moreno-Cadenas, "Spiking neural network approaches PCA with metaheuristics", *Electron. Lett.*, vol. 56, no. 10, pp. 488–490, 2020.