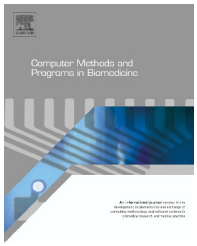
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE XXX (2013) XXX–XXX

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Design of fuzzy classifier for diabetes disease using Modified Artificial Bee Colony algorithm

Fayssal Beloufa, M.A. Chikh*

Biomedical Engineering Laboratory, Tlemcen University, Algeria

ARTICLE INFO

Article history:
Received 1 April 2012
Received in revised form
15 July 2013
Accepted 17 July 2013

Keywords:
Artificial Bee Colony
Interpretable classification
Fuzzy rules
Diabetes disease

ABSTRACT

In this study, diagnosis of diabetes disease, which is one of the most important diseases, is conducted with artificial intelligence techniques. We have proposed a novel Artificial Bee Colony (ABC) algorithm in which a mutation operator is added to an Artificial Bee Colony for improving its performance. When the current best solution cannot be updated, a blended crossover operator (BLX-α) of genetic algorithm is applied, in order to enhance the diversity of ABC, without compromising with the solution quality. This modified version of ABC is used as a new tool to create and optimize automatically the membership functions and rules base directly from data. We take the diabetes dataset used in our work from the UCI machine learning repository. The performances of the proposed method are evaluated through classification rate, sensitivity and specificity values using 10-fold cross-validation method. The obtained classification rate of our method is 84.21% and it is very promising when compared with the previous research in the literature for the same problem.

1. Introduction

Diabetes is a complex and complicated disease characterized by either lack of insulin or a resistance to insulin, a hormone which is crucial for metabolism of blood sugar. In a healthy person, the pancreas produces insulin to help metabolize sugar in the blood and maintain blood glucose (sugar) levels within their normal range. Diabetics are unable to produce insulin or are resistant to insulin, and consequently cannot remove glucose from the bloodstream. Whether there is inadequate insulin or insulin resistance, glucose levels in the blood increase and cause severe health problems. The classification of diabetes includes two main clinical classes: Type 1 diabetes (previously called "juvenile diabetes") is an autoimmune disorder in which the insulin producing beta cells are destroyed by the body's immune system. As a result the body is unable to produce insulin, Type 2 diabetes (previously called "adult onset diabetes"); in this type insulin is produced in insufficient amounts and/or cannot be used by the body to control blood sugar levels [1]. Therefore the need to detect and treat diabetes becomes obvious to reduce its incidence and costly associated metabolic disease. For this reason, in recent times, many machine learning techniques have been considered to design automatic diagnosis system for diabetes. This paper specifically focuses on the use of fuzzy modeling method to detect medical problems which relies on discovering human comprehensible knowledge. Fuzzy logic is originally proposed by Zadeh [2], who aimed to improve a classification and decision support systems by using fuzzy sets to define the overlapping class definitions. The application of fuzzy "if-then" rules also improves the interpretability of the results and provides more insight cognition into the classifier structure and the decision making process [3]. The performance of fuzzy classifier system depends on the "if-then" rules and their numbers that are generated from numerical data or human experiences. More

* *Corresponding author*. Tel.: +213 670160191.
E-mail addresses: beloufa.fayssal@gmail.com (F. Beloufa), mea_chikh@mail.univ-tlemcen.dz (M.A. Chikh).
0169-2607/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cmpb.2013.07.009

rules might enhance the classification result but increase the computation load. Many intelligent methods, such as heuristic approaches [4], neural networks [5–7] and clustering method [8–10] have been proposed for building optimal fuzzy classifiers but in recent years Evolutionary Algorithms (EAs) have been widely used to optimize fuzzy classifiers. In the literature several EAs like Genetic Algorithm (GA), particle swarm optimization (PSO) and ant colony optimization (ACO) have been proposed to produce fuzzy classification system. Ganji et al. [11] used ACO algorithm to generate fuzzy classification rules called FSC-ANTMINER. They have used artificial ants in order to explore the search space and gradually make candidate fuzzy rules. Other works have adopted GA to optimize fuzzy classifiers [12–14]. In [15] Sanie et al. used a hybrid genetic algorithm to produce fuzzy rules and boosted it with an Ant Colony Optimization (ACO) heuristic based on local search to improve the quality of their final classification system. In [16] the authors proposed a fuzzy modeling framework able to generate automatically a rule base through a two stage genetic based search. On the other hand, PSO based approaches have proved their efficiency when optimizing fuzzy classification system. In [17] Sousa et al. proposed a first application of PSO as a new tool for classification rule discovery. Holden and Freitas [18] used a hybrid Particle Swarm Optimization and Ant Colony Optimization algorithm for discovering classification rules in data mining. Rani and Deepa [19] proposed a particle swarm optimization approach for optimal design of fuzzy classifier called PSOFLC. Experiments are performed on IRIS dataset, where the proposed technique is compared to two well-known classification techniques, including Genetic Fuzzy Classifier and Gaussian Fuzzy Classifier. The framework used for designing the fuzzy model from the available input–output data through PSO algorithm was proposed in [20] and the detailed encoding method was also provided. Khosla et al. [21] compared the computational efforts of PSO and GA using the similar method in [20]. Recently, a new Artificial Bee Colony algorithm (ABC) [22] was proposed as an alternative of the traditional Evolutionary Algorithms. The ABC algorithm is very simple and flexible when compared to other swarm based algorithms such as Particle swarm optimization. It does not require external parameters like mutation and crossover rates, which are hard to be defined in prior. The algorithm combines local search method with global one and tries to reach a certain balance between exploration and exploitation [22]. The ABC algorithm has been successfully applied to a wide range of applications such as clustering analysis [23], neural network training [24], function optimization [25,26], Wireless Sensor Network [36,37] and Symbolic Regression [38].

In this paper, some modifications of a standard Artificial Bee Colony are introduced. The modified ABC differs from the original one in the sense that if there is no improvement in fitness function, the positions are mutated using blended crossover operator of genetic algorithm for more exploitation and exploration. This modified version of Artificial Bee Colony is proposed as a new tool for building a compact fuzzy rule based classifier without any a priori knowledge. According to our experiments, our model produces compact fuzzy rules based classifier and can work efficiently for diabetes diseases. In addition, the results obtained in this paper show that the proposed technique outperforms other methods in term of classification rate. The rest of this paper is organized as follows: Section 2 introduces the related works on diabetes disease diagnosis. In Section 3 the Fuzzy rule based classification model, Artificial Bee Colony algorithm are described. We discuss about our proposed method in detail in Section 4. The results are presented and discussed in Section 5, and finally, in Section 6, we conclude the paper.

2. Related work

Much research efforts have been performed till today to examine and classify diabetes usually based on artificial intelligence techniques in literature. In [27] a generalized discriminant analysis and least square support vector machine was used for diagnosing Pima Indian diabetes. In this work the authors have reported 79.16% classification accuracy. Polat and Gunes [28] used attribute weighted artificial immune system with 10-fold cross validation method. They obtained a classification accuracy of 75.87%. Tand and Tseng [29] developed GA-based methods to estimate a weight vector of the feature vector applied in the fuzzy k-NN estimation. The accuracy was 81.6% using Binary-coded Genetic Algorithms (BGA) and 82% using Real-coded Genetic Algorithms (RGA) with 3 features. Polat and Günes [30] obtained 89.47% classification accuracies using a principal component analysis and an adaptive neuro-fuzzy inference in diabetes disease diagnostic. In Temurtas et al. [31] a multi-layer neural network and a Probabilistic Neural Network (PNN) were used for diagnosing Pima Indian diabetes. They have reported respectively 79.62% and 78.05% in terms of correct classification rate.

3. Theory

3.1. Fuzzy rules based classification model

A fuzzy classifier consists of linguistic rules which are easy to interpret by the user. The classifier is not a black box; it can be checked for plausibility. This is very important for decision support systems, where users do not accept a computer's evaluation, unless they understand why and how a certain recommendation was given. Therefore, the fuzzy classifier can be a convenient tool in the diagnostic process. The classification problem consists of m training patterns $X_p(x_{p1}, \ldots, x_{pn})(p = 1, 2, \ldots, m)$ from M classes where x_{pi} is the ith attribute value $(i = 1, 2, \ldots, n)$ of the pth training pattern. In this study the fuzzy "if-then" rules can be expressed as follows:

$$R_j : \text{if} x_{p1} \text{ is} A_{j1} \text{ and} x_{p2} \text{ is} A_{j2} \text{ and} \ldots \text{and} x_{pn} \text{ is} A_{jn}$$
$$\text{THEN} x_p(x_{p1}, \ldots, x_{pn}) \text{ belongs to class C} \quad j = 1, \ldots, N \tag{1}$$

Here R_j is the label of the jth rule, $X = (x_1, \ldots, x_n)$ is an n-dimensional pattern vector, A_{ji} is an antecedent fuzzy set and C $(c = 1, 2, \ldots, M)$ is a class label. We use the fuzzy reasoning method of the winning rule (classical approach) [32] for classifying new patterns by the rules base. The single winner rule R_w is determined for a new pattern $x_p(x_{p1}, \ldots, x_{pn})$ as:

$$w = \text{argmax}\{\mu_j(x_p)\} | j = 1, 2, \ldots, N| \tag{2}$$

Please cite this article in press as: F. Beloufa, M.A. Chikh, Design of fuzzy classifier for diabetes disease using Modified Artificial Bee Colony algorithm, Comput. Methods Programs Biomed. (2013), http://dx.doi.org/10.1016/j.cmpb.2013.07.009