COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 121 (2015) 137–148

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

# A Swarm Optimization approach for clinical knowledge mining

J. Jabez Christopher[a], H. Khanna Nehemiah[a,*], A. Kannan[b]

[a] Ramanujan Computing Centre, Anna University, Chennai 600025, Tamil Nadu, India
[b] Department of Information Science and Technology, Anna University, Chennai 600025, Tamil Nadu, India

ARTICLE INFO

*Article history:*
Received 9 December 2014
Received in revised form
27 May 2015
Accepted 28 May 2015

*Keywords:*
Bio-inspired computing
Swarm intelligence
Clinical decision making
Medical decision support systems
Rule-based classification

ABSTRACT

*Background and objectives:* Rule-based classification is a typical data mining task that is being used in several medical diagnosis and decision support systems. The rules stored in the rule base have an impact on classification efficiency. Rule sets that are extracted with data mining tools and techniques are optimized using heuristic or meta-heuristic approaches in order to improve the quality of the rule base. In this work, a meta-heuristic approach called Wind-driven Swarm Optimization (WSO) is used. The uniqueness of this work lies in the biological inspiration that underlies the algorithm.
*Methods:* WSO uses Jval, a new metric, to evaluate the efficiency of a rule-based classifier. Rules are extracted from decision trees. WSO is used to obtain different permutations and combinations of rules whereby the optimal ruleset that satisfies the requirement of the developer is used for predicting the test data. The performance of various extensions of decision trees, namely, RIPPER, PART, FURIA and Decision Tables are analyzed. The efficiency of WSO is also compared with the traditional Particle Swarm Optimization.
*Results:* Experiments were carried out with six benchmark medical datasets. The traditional C4.5 algorithm yields 62.89% accuracy with 43 rules for liver disorders dataset where as WSO yields 64.60% with 19 rules. For Heart disease dataset, C4.5 is 68.64% accurate with 98 rules where as WSO is 77.8% accurate with 34 rules. The normalized standard deviation for accuracy of PSO and WSO are 0.5921 and 0.5846 respectively.
*Conclusion:* WSO provides accurate and concise rulesets. PSO yields results similar to that of WSO but the novelty of WSO lies in its biological motivation and it is customization for rule base optimization. The trade-off between the prediction accuracy and the size of the rule base is optimized during the design and development of rule-based clinical decision support system. The efficiency of a decision support system relies on the content of the rule base and classification accuracy.



## 1. Introduction

Clinical decision making is a pervasive task that needs to be accurate and quick. Medical diagnosis, prognosis, patient-care and research have many objectives and constraints, in the midst of which optimal decisions have to be made. In order to provide quality health care, the process of decision making is enhanced and supported by various systems such as Decision Support Systems (DSS) [1], Rule-Based Clinical Decision

* *Corresponding author*. Tel.: +91 044 22358013; mobile: +91 9381038339.
E-mail address: nehemiah@annauniv.edu (H.K. Nehemiah).

http://dx.doi.org/10.1016/j.cmpb.2015.05.007
0169-2607/

Support Systems [2], Expert System [3], Automatic Diagnosis System [4], Diagnosis Support System [5], Clinical Information System [6], Medical-Record System [7] and hybrid intelligent system [8].

Clinical DSS focus on providing information and knowledge for decision makers at various levels in a health care organization. Every day increasing amount of health-care data is maintained in repositories but the emphasis is on mining knowledge from clinical data. The extracted knowledge should be novel, interesting, precise and comprehensible so as to improve the decision making process. Efficient and robust computational algorithms, are required to develop an optimized decision making model.

Several mathematical, evolutionary and swarm intelligence optimization approaches have been proposed in the literature. Some of these approaches used in medical informatics and clinical diagnosis are the following: Numerical analysis [9], Linear programming [10], Non-Linear analysis [11], genetic algorithm (GA) [12], Immune-GA [13], Simulated Annealing [14], Tabu Search [15], Ant Colony Optimization [16], Artificial Bee Colony (ABC) [17] Algorithm, Particle Swarm Optimization (PSO) [18,19].

Harmony search is a metaheuristic approach for optimization of continuous variables [20]. The approach is inspired from the music and harmony construction. Each musician can be compared to a decision variable and the pitches are possible values of that variable. There are three rules followed by the variables to choose a feasible value which are as follows: first, choosing a value from the harmony search (HS) memory, second, choosing an adjacent value from the existing set of values and third choosing a random value. A certain set of values for each variable, which maximizes or minimizes the objective, constitutes the optimal solution.

Erol et al. proposed an optimization method inspired by the evolution of the universe [21]. The big bang and the big crunch theory consist of two phases namely, the big bang phase and a big crunch phase. The big bang phase deals with the creation (initialization) of a population in the search space. The entire set of solutions initialized in the big bang phase is given as input to the big crunch phase, and the centre of mass ($x_c$) is obtained as output. The centre of mass is considered to be the optimal solution or the best-fit individual.

Oftadeh et al. proposed an optimization algorithm based on the group hunting behaviour of animals [22]. Animals that hunt their prey in groups follow certain procedures to effectively get hold of the prey. The group leader is closer to the prey and hence the members of the group follow the leader. The group leader can be compared to an optimal solution and the other hunting animals as feasible (candidate) solutions.

Yang proposed the Firefly Algorithm (FA) for global optimization problems [23]. Fireflies emit light to attract the prey. The flashing light is formulated in such a way that it is associated with the objective function to be optimized. In optimization, each fly is analogous to a solution and the fly with the maximum intensity of brightness is considered as an optimal solution.

Particle Swarm Optimization [24], a population based meta-heuristic stochastic optimization algorithm, has been applied for optimizing typical data mining tasks such as classification, regression, clustering and association analysis. An overview of PSO is presented in the related works section.

Due to several drawbacks and complexities in the design and application of metaheuristic approaches, new algorithms and approaches are developed by the research community. In this work, a swarm intelligence-based metaheuristic optimization approach is used to build efficient rule-based clinical decision support systems.

The rest of the paper is organized as follows: Section 2 presents the related work. Section 3 presents the biological motivation of the proposed optimization approach. In Section 4, a description about the datasets and the application of WSO for clinical rule-based classification is discussed. Section 5 presents the results, findings and suggestions along with the pros and cons of different approaches used. Finally Section 6 concludes the paper highlighting the scope for future work.

## 2. Related work

This section provides an overview about the generic global PSO, followed by its variants and modifications proposed by different researchers. Application of PSO for clinical data mining is presented. Similar optimization approaches such as Glowworm Swarm Optimization (GSO), Wind Driven Optimization (WDO) are also discussed. Finally, the issues and drawbacks of swarm intelligence based optimizers are discussed and the significance of this work is highlighted.

### 2.1. *Particle Swarm Optimization*

Kennedy and Eberhart proposed the original Particle Swarm Optimization algorithm in the year 1995 [24], after which various PSO variants have been developed to improve its performance. For many cases, the variations and changes have been imposed over the methods that control the working of the algorithm. In short, these changes range from adding constants in the particles' velocity update rule to stand-alone algorithms that are used as extensions of hybrid PSO algorithms. In this section we present the generic PSO algorithm, different variations in PSO and the modifications in the variations.

A population based stochastic optimization algorithm has a swarm of particles (feasible solutions) moving towards an optimal position (optimal solution) in a search space. The basic position update and velocity update rules in the generic PSO are as follows:

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (1)$$

$$v_{ij}(t+1) = v_{ij}(t) + c_1 r_{1j}(t)[y_{ij}(t) - x_{ij}(t)] + c_2 r_{2j}(t)[\hat{y}_j(t) - x_{ij}(t)] \quad (2)$$

where $x_{ij}(t)$ is the position of particle $i$ in dimension $j$ at step $t$, $v_{ij}(t)$ is the velocity of particle $i$ in dimension $j = 1,2,\ldots,n_x$ at time $t$. $c_1$ *and* $c_2$ are cognitive acceleration coefficient and social acceleration coefficient respectively. The variables $r_1$, $r_2$