



# A cooperative swarm intelligence algorithm based on quantum-inspired and rough sets for feature selection



Djaafar Zouache<sup>a</sup>, Fouad Ben Abdelaziz<sup>b,\*</sup>

<sup>a</sup> Computer Science Department, Mohamed El Bachir El Ibrahimi University, Bordj Bou Arreridj, Algeria

<sup>b</sup> NEOMA Business School, Rouen Campus, France

## ARTICLE INFO

### Keywords:

Feature selection  
Classification  
Rough sets  
Quantum computation  
Swarm intelligence

## ABSTRACT

Feature selection is an important preprocessing step for classification as it improves the accuracy and overcomes the complexity of the classification process. However, in order to find a potentially optimal feature subset for the feature selection problem, it is necessary to design an efficient exploration approach that can explore an enormous number of possible feature subsets. It is also necessary to use a powerful evaluation approach to assess the relevance of these feature subsets. This paper presents a new cooperative swarm intelligence algorithm for feature selection based on quantum computation and a combination of Firefly Algorithm (FA) and Particle Swarm Optimization (PSO).

Quantum computation ensures a good trade-off between the exploration and the exploitation of the search space while the combination of the FA and PSO enables an effective exploration of all the possible feature subsets. We use rough set theory to assess the relevance of the potential generated feature subsets. We tested the proposed algorithm on eleven UCI datasets and compared with a deterministic rough set reduction algorithms and other swarm intelligence algorithms. The experiment results show clearly that our algorithm provides a better rate of feature reduction and a high accuracy classification.

## 1. Introduction and literature reviews

Classification poses two major problems. The first is the great number of condition features, which may lead to highly complex computation in the classification model. This is the case, for instance, in classification using neural networks, where the obtained model contains a huge nodes number. The second problem is related to the presence of features that are redundant and/or not sufficiently relevant to the decision feature, which reduces the quality and performance of the classification algorithm.

Feature selection, as it addresses an exponential number of solutions and is considered as a necessary preliminary step to classification, is proven to be an NP complete problem by Davies and Russell (1994).

The results of feature selection make it possible to reduce the complexity of constructing a classification model, and in many cases, it increases the performance of classification. Feature selection addresses two important questions (i) how to explore the different possible subsets of features to generate the best one; and (ii) how to evaluate the relevance of possible feature subsets?

Depending on the exploration and evaluation strategies, feature selection methods proposed in literature may be divided into two types:

filter feature selection methods and wrapper feature selection methods (Langley, 1994). The main difference between the two types is that wrapper feature selection uses a classification algorithm to assess the relevance of a feature subset during the feature selection process. However, in filter feature selection, the relevance of feature subsets is assessed independently of the classification algorithm. Generally, wrapper feature selection provides greater quality than filter feature selection. However, wrapper feature selection methods are computationally expensive (Dash & Liu, 1997).

Several theories are used to assess the relevance of the generated feature subset, such as: mutual information (Estévez, Tesmer, Perez, & Zurada, 2009), rough sets (Pawlak, 1982), fuzzy rough sets (Wygalak, 1989) and the Dempster-Shafer theory (Dempster, 1967). Rough set theory, introduced in 1982 by Pawlak (1982), is considered a powerful tool for feature selection, association rule mining and knowledge discovery from categorical data (Degang, Changzhong, & Qinghua, 2007; Pawlak & Skowron, 2007b; Slowinski & Vanderpooten, 2000; Swiniarski & Skowron, 2003). Rough set theory deals with uncertain environments like feature selection in the presence of heterogeneous, incomplete information systems. Heterogeneity means that the information system contains both categorical and continuous features.

\* Corresponding author.

E-mail addresses: [djaafarzouache@yahoo.fr](mailto:djaafarzouache@yahoo.fr) (D. Zouache), [Fouad.BEN.ABDELAZIZ@neoma-bs.fr](mailto:Fouad.BEN.ABDELAZIZ@neoma-bs.fr) (F. Ben Abdelaziz).

Incompleteness means that some features of the information system contain missing or null values.

Two methods of rough set theory for feature selection in a heterogeneous, incomplete information system are proposed in the literature. The first method pre-treats the information system, including the discretization of continuous features, and the detection of missing values before application of rough set theory concepts. The second method adapts the concepts of the theory and applies them directly to an incomplete heterogeneous information system.

To adapt rough set theory to heterogeneous feature selection, some researchers have proposed methods based on positive region. These methods include the neighborhood rough set model (Hu, Yu, & Xie, 2008b) and the  $k$ -nearest neighbor rough set (Hu, Liu, & Yu, 2008a). However, these methods have high computational requirements, and cannot handle large information systems (Hu et al., 2008a).

Other researchers have proposed to integrate the concepts of fuzzy logic in the rough set theory (Fuzzy rough set) to effectively deal with mixed data. These proposals include: fuzzy-rough quick reduct algorithm (Jensen & Shen, 2007), fuzzy-rough model and information granulation (Hu, Xie, & Yu, 2007), and attributes reduction using fuzzy rough sets (Tsang, Chen, Yeung, Wang, & Lee, 2008). Again, these methods are computationally demanding and do not generate fuzzy relationships effectively.

In recent years, several researchers have tried to adapt rough set theory concepts to deal with feature selection in an incomplete information system as the methods that are based on changes in the discernibility relationship to better support the missing values: Rough sets for feature selection in incomplete information systems (Kryszkiewicz, 1998), tolerance relation-based rough sets (Meng & Shi, 2009), approximation reduction in inconsistent incomplete decision tables (Qian, Liang, Li, Wang, & Ma, 2010), covering rough sets (Degang et al., 2007), attribute reduction algorithm based on conditional entropy under incomplete information system (Teng, Zhou, Sun, & Li, 2010), and other methods.

Several metaheuristics have been proposed to solve the feature selection problem, such as genetic algorithms (Davis, 1991), particle swarm optimization (PSO) (Kennedy, 2011) and ant colony optimization (ACO) (Dorigo & Blum, 2005). Considerable efforts have been made to hybridize metaheuristics and rough set theory for feature selection, such as genetic algorithm using rough set theory (Zhai, Khoo, & Fok, 2002), particle swarm optimization based on rough sets (Wang, Yang, Teng, Xia, & Jensen, 2007; Inbarani, Azar, & Jothi, 2014; Bae, Yeh, Chung, & Liu, 2010) and ant colony optimization with rough sets (Chen, Miao, & Wang, 2010; Jensen & Shen, 2005; Ke, Feng, & Ren, 2008).

Recently, many swarm intelligence algorithms have been proposed using rough set theory for the feature selection: fish swarm algorithm based on rough sets (Chen, Zhu, & Xu, 2015), new rough set attribute reduction algorithm based on grey wolf optimization (Yamany, Emary, & Hassanien, 2016a), attribute reduction algorithm based on rough set and improved artificial fish swarm algorithm (Luan, Li, & Liu, 2016), neighborhood rough set reduction with fish swarm algorithm (Chen, Zeng, & Lu, 2016) and attribute reduction using rough sets and flower pollination optimization (Yamany, Emary, Hassanien, Schaefer, & Zhu, 2016b). However, these methods do not address the case of incomplete heterogeneous information systems.

In this paper, we use swarm intelligence algorithm tools such as parallelism, decentralization, and cooperation to solve the feature selection problem. We will explore the search space by hybridizing two swarm metaheuristics (FA and PSO) and integrate basic concepts of quantum computation, such as quantum measurement and qubits superposition state, into two-swarm intelligence algorithms to effectively diversify search space. By using the concepts of quantum computation, one can break the complexity of the feature selection problem where the superposition of qubits provides the opportunity for the quantum register to contain multiple possible feature subsets simultaneously. We

will use rough set theory to assess the relevance of a feature subset in an information system without preprocessing.

The rest of this paper is organized as follows. We introduce the main concepts in Section 2, including a short presentation of particle swarm optimization (Section 2.2), the firefly algorithm (Section 2.3), quantum computation (2.4), and the preliminary concepts of rough set theory (Section 2.1). Section 3 provides detailed procedures of our cooperative swarm intelligence algorithm for feature selection based on quantum-inspired and rough sets (QCSIA-FS). In Section 4, we present the results of various numerical experiments carried out for feature selection. The conclusion, in Section 5, suggests some directions for future research.

## 2. Background

### 2.1. Preliminary concepts of rough sets

This section introduces some concepts of rough sets theory in the context of information systems that we will use in our algorithm to evaluate candidate feature subsets. These concepts are the information system, indiscernibility, set approximation, neighborhood rough sets, and feature dependency (Pawlak & Skowron, 2007a).

#### 1. Information system

An information system is defined by the tuple  $(U, A = C \cup D, V, f_a)$ , where  $U$  is a finite non-empty set of objects,  $C$  is a non-empty set of features called the set of condition features,  $D$  is a non-empty set of features called the set of decision features and  $C \cap D = \emptyset; V = \cup_{a \in A} V_a$ , where  $V_a$  is the set of values of the feature domain  $a \in A$ , and  $f_a: U \rightarrow V_a$  is an information function defined from  $U$  towards  $V_a$ .

#### 2. Indiscernibility

For every condition feature subset  $B \subset C$ , there is an associated equivalence relation defined by:

$$IND(B) = \{(x, y) \in U^2 \mid \forall a \in B, f_a(x) = f_a(y)\} \quad (1)$$

$IND(B)$  is the  $B$ -indiscernibility relation. This relation means that couples of objects  $(x, y) \in U^2$  are indiscernible by the set of features  $B$ . The relation  $IND(B)$  generates a partition  $U/IND(B) = \{[x]_B \mid x \in U\}$  over  $U$ , where  $[x]_B$  is the equivalence class of an object which consists of all objects  $y \in U$  such that  $x$  is indiscernible with  $y$  by the features set  $B$ .

#### 3. Set approximation

Let  $B \subset C$  be the set of condition features,  $[x]_B$  be the equivalence class of each object by the feature subset  $B$ . The approximation of the set of object  $X \subset U$  by using the equivalence class  $[x]_B$  is given by the lower  $\underline{B}X$  and the upper approximation  $\overline{B}X$ . The lower approximation of  $X$  is defined by:

$$\underline{B}X = \{x \in U \mid [x]_B \subseteq X\} \quad (2)$$

The upper approximation of  $X$  is defined by:

$$\overline{B}X = \{x \in U \mid [x]_B \cap X \neq \emptyset\} \quad (3)$$

The lower approximation of  $X$  is called the positive region of  $X$  and is denoted by  $POS_B(X)$ .

#### 4. Dependency of features

Let  $I = (U, A = C \cup D, V, f_a)$  be an information system. The partitioning of the universe  $U$  by the indiscernibility relation of the decision feature  $D$  is:

$$U/IND(D) = \{D_1, D_2, \dots, D_k\} \quad (4)$$

with  $U = \cup_{i \in \{1, \dots, k\}} D_i$ ,  $\underline{B}D_i$  is the lower approximation of each partition  $D_i$  by the set of condition features  $B$ . The positive region of the decision feature  $D$  which respects the set of condition features  $B$ , denoted by  $POS_B(D)$  is given by:

$$POS_B(D) = \cup_k^i POS_B(D_i) \quad (5)$$

Download English Version:

<https://daneshyari.com/en/article/7541514>

Download Persian Version:

<https://daneshyari.com/article/7541514>

[Daneshyari.com](https://daneshyari.com)