Quantum-inspired evolutionary approach for selection of optimal parameters of fuzzy clustering

Article *in* International Journal of Systems Assurance Engineering and Management · November 2017

DOI: 10.1007/s13198-017-0681-x

3 authors, including:

Neha Bharill
Indian Institute of Information Technology Dharwad

op Patel
Indian Institute of Technology Indore

Int J Syst Assur Eng Manag
DOI 10.1007/s13198-017-0681-x

ORIGINAL ARTICLE

Quantum-inspired evolutionary approach for selection of optimal parameters of fuzzy clustering

Neha Bharill[1] · Om Prakash Patel[1] · Aruna Tiwari[1]

Revised: 22 June 2017

Abstract Recently, Fuzzy c-Means (FCM) algorithm is most widely used because of its efficiency and simplicity. However, FCM is sensitive to the initialization of fuzziness factor (m) and the number of clusters (c) due to which it easily trapped in local optima. A selection of these parameters is a critical issue because an adverse selection can blur the clusters in the data. In the available fuzzy clustering literature, cluster validity index is used to determine the optimal number of clusters for the dataset, but these indexes may trap into the local optima due to the random selection of m. From the perspective of handling local optima problem, we proposed a hybrid fuzzy clustering approach referred as quantum-inspired evolutionary fuzzy c-means algorithm. In the proposed approach, we integrate the concept of quantum computing with FCM to evolve the parameter m in several generations. The evolution of fuzziness factor (m) with the quantum concept aims to provide the better characteristic of population diversity and large search space to find the global optimal value of m and its corresponding value of c. Experiments using three real-world datasets are reported and discussed. The results of the proposed approach are compared to those obtained from validity indexes like V_{CWB} and V_{OS} and evolutionary fuzzy based clustering algorithms. The results show that proposed method achieves the global optimal value of m, c with a minimum value of fitness function and shows significant improvement in the convergence times (the number of iterations) as compared to the state-of-the-art methods.

Keywords Quantum computing · Fuzzy clustering · Fuzzy c-Means · Cluster validity index

✉ Neha Bharill
phd12120103@iiti.ac.in

Om Prakash Patel
phd1301201003@iiti.ac.in

Aruna Tiwari
artiwari@iiti.ac.in

[1] Department of Computer Science and Engineering, Indian Institute of Technology Indore, Simrol, Indore, Madhya Pradesh, India

1 Introduction

Clustering is an unsupervised learning approach which plays a significant role in pattern recognition (Webb 2003). It is the process of grouping the data points in the clusters (groups) so that the data points belong to the same cluster should exhibit the similarity with each other than those which are in different groups (Jain 2010; Kaufman and Rousseeuw 2009). Clustering algorithms are broadly classified into two main groups: hard clustering and fuzzy clustering methods (Xu and Wunsch 2005). Hard clustering algorithm partition the datasets into clusters, where each object is allowed to belong to a single cluster. Fuzzy clustering algorithms allow each object to belong to multiple clusters with varying membership degree which represents how far the object belongs to the cluster.

Fuzzy c-Means is one of the most widely used fuzzy clustering methods proposed by Bezdek et al. (1984). It is widely applied to the datasets that have highly overlapping groups. Since FCM is an efficient algorithm because it generates satisfactory results in many applications. However, FCM suffers from various shortcomings such as determining the most proper number of clusters (c), and the

Published online: 04 November 2017

fuzziness factor (m) is an important issue in fuzzy cluster analysis, which significantly affects the clustering results of FCM (Baskir and Turksen 2013; Stetco et al. 2015). But, due to the scarcity of computationally fast methods for determining the value of these parameters, random selection of these parameters makes the iterative process falling into the local optimal solution and may lead to the inclusion of purely random fluctuations in the results by ignoring potentially important data (Izakian and Abraham 2011). In the literature, the most commonly used approach for finding the optimal number of clusters (c) and the fuzziness factor (m) is by predefined criteria function known as a cluster validity index (CVI). Though CVIs determine the optimal number of clusters, it also assists in selecting the optimal value of fuzziness factor in the FCM. The FCM algorithm is executed several times with different values of m, and the optimal value of m is determined when CVIs attain their extreme values. Some of these validity indexes are Bezdek (1973, 1974, 1998), Fukuyama and Sugeno (1989), Rezaee et al. (1998), Kim et al. (2004), Bharill and Tiwari (2014), Zhou et al. (2014), Erilli et al. (2011).

In addition to this, there have been lots of studies carried out for the selection of the fuzziness factor in the FCM. Pal and Bezdek (1995) investigated that fuzziness factor (m) is an important factor influencing the effectiveness of FCM and present a heuristic rule suggesting that an appropriate value of m is limited to [1.5, 2.5]. They assumed that median value $m = 2$ could be chosen as the optimal value without any special constraints. Yu et al. (2004), present a theoretical upper bound for m that can prevent the sample mean from being a unique optimizer of the FCM objective. Furthermore, Ozkan and Turksen (2004) proposed a method that determines the upper and lower bound value of m is 1.4 and 2.6 respectively, based on the expansion of membership function around cluster center and mass center. Wu (2012) presented a new guideline for selecting the parameter m based on the robust analysis of FCM. Huang et al. (2012) studied the relationship between m and the distribution of clusters and identified a range of values of m based on the behavior of membership function on two particular data points. However, there is no theoretical guidance for the selection of m in the FCM.

Furthermore, regarding the selection of an appropriate number of clusters and solving the problem of local convergence in FCM, many evolutionary hybrid data clustering algorithms is proposed. Li et al. (2007) proposed a PSO-based fuzzy clustering algorithm using the advantage of global search capabilities of PSO algorithm to overcome the shortcoming of FCM. Izakian and Abraham (2011) proposed a hybrid fuzzy data clustering algorithm combining the merits of FCM and fuzzy PSO to overcome the local convergence problem in the FCM. Bandyopadhyay and Maulik (2001) proposed an evolutionary approach based on a variable length genetic algorithm for finding the number of clusters. Recently, the work of merging quantum computing and evolutionary computing has simulated the studies of quantum-inspired evolutionary algorithms and their applications. Han and Kim (2002) proposed a quantum-inspired evolutionary algorithm (QEA) to solve the combinational problem. QEA used a Q-bit as a probabilistic representation and a Q-bit individual consist of a string of a Q-bit. The Q-bit representation in the quantum computing represents the linear superposition of states probabilistically which provides the better characteristic of population diversity than other representations. This enables the more effective exploration of the search space, leading to the global optimum solution in the search space by effectively eliminating premature convergence. Inspired by the idea Hung et al. (2013) proposed a quantum-modeled Fuzzy c-Means algorithm for remotely sensed multiband image segmentation. It uses this model for providing diversity in selecting the initial fuzzy clustering membership as an input to the FCM clustering and thus produces better results than a traditional FCM algorithm. Wang et al. (2014) proposed an approach for determining the multi-distribution center location by merging real-parameter quantum-inspired evolutionary algorithm (RQIEA) and FCM. It overcomes the local search defect of the FCM by making the optimization results independent of the choice of initial values of the cluster centroid (Pal and Bezdek 1995; Yu et al. 2004). Xiao et al. (2010) proposed a quantum-inspired genetic algorithm for k-means clustering (KMQGA) which uses quantum bit representation as well as the typical genetic algorithm operations (selection, crossover, and mutation to determine the optimal number of clusters and to provide the optimal cluster centroids. Silva Filho et al. (2015) proposed hybrid methods for fuzzy clustering refers to as FCM–IDPSO and FCM2–IDPSO which combines the FCM with the recent version of PSO. In this, the IDPSO adjust PSO parameters dynamically during execution, aiming to provide a better balance between exploration and exploitation, avoiding falling into local minima quickly and thereby obtaining better solutions.

In this paper, a hybrid fuzzy clustering algorithm is proposed, which is named as Quantum-Inspired Evolutionary Fuzzy c-Means (QIE–FCM) algorithm. The proposed approach uses the merits of quantum computing for finding the global optimal value of m and its corresponding value of c in the FCM. QIE–FCM uses a quantum bit individual to represent the fuzziness factor (m) in each generation (g, user defined number). Before passing the fuzziness factor (m) for the clustering of data using the FCM algorithm, the quantum bits of fuzziness factor have to be encoded into the binary representation to explore the quantum search space, and after that, it has to be converted into real coded representation. For the real coded value of m obtained in generation g, several iterations of the FCM algorithm are executed by varying values of c in the range of $[c_{\min},..., c_{\max}]$ where $c_{\min} = 2$ and $c_{\max} = \sqrt{(n)}$ (n is the number of instances) (Höppner 1999). During each generation (g) of QIE–FCM, a cluster validity index is required to validate the obtained fuzzy partitions corresponding to different values of c, so this paper uses VI_{DSO} (Bharill and Tiwari 2014) index for the cluster validity. In each generation of the QIE–FCM, the quantum bits of fuzziness factor is updated using the quantum rotational gate (Han and Kim 2004). After several generations of QIE–FCM, the proposed algorithm guarantees to find the global optimum value of fuzziness factor (m) and its corresponding number of clusters (c) by effectively eliminating premature convergence in the FCM. The proposed method improves the way of initialization of the fuzziness factor (m) in the FCM and provides the diversity in selecting the optimal value of m and c from a large quantum search space. Real life datasets (Lichman 2015) are used to measure the performance of the proposed approach in comparison with standard validity indexes. The performance is measured in terms of faster convergence times (the number of iterations), lower value of the fitness function, the global optimal value of fuzziness factor (m) and the number of clusters (c). Furthermore, proposed approach shows the superior clustering results in comparison with evolutionary fuzzy based clustering algorithms (Bandyopadhyay and Maulik 2001; Hung et al. 2013; Wang et al. 2014; Xiao et al. 2010; Silva Filho et al. 2015; Bandyopadhyay 2011) in terms of the low value of fitness function and the optimal number of clusters.

The rest of the paper is structured as follows: Section 2 briefly introduces the concept of quantum computing. Section 3 introduces fuzzy c-means clustering. Section 4 summarizes fuzzy cluster validity indexes; Sect. 5 present the description of proposed QIE–FCM algorithm. Experimental results and detailed analysis are presented in Sect. 6; Sect. 7 discusses the findings and present future work.

2 Quantum computing

Before describing QIE–FCM, we present the introduction of quantum computing briefly. In quantum computers, the smallest unit of information representation is called quantum bits or (Q-bit) (Hey 1999). A Q-bit is fundamentally different from binary bits used in traditional digital computers for data representation. A single binary bit can represent only two-state, i.e. "0" and "1". However, the Q-bit may be in the "0" state, "1" or in any superposition of the two states. The state of a Q-bit can be represented by the formula (1):

$$Q = \alpha \mid 0\rangle + \beta \mid 1\rangle = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{1}$$

where α and β are the complex numbers representing the probabilities that Q-bit may appear in "0" state and "1" state. Here, α^2 represent the probability of a Q-bit in "0" state, whereas β^2 denote the probability of a Q-bit in "1" state, which is defined as follows:

$$\alpha^2 + \beta^2 = 1; \quad 0 \le \alpha \le 1, 0 \le \beta \le 1 \tag{2}$$

The advantage of quantum concept is that a Q-bit individual consist of n number of Q-bits, which is defined as follows:

$$Q = \left[\begin{array}{c|c|c|c} \alpha_1 & \alpha_2 & \ldots & \alpha_n \\ \beta_1 & \beta_2 & \ldots & \beta_n \end{array} \right] \tag{3}$$

Suppose, a Q-bit individual with three Q-bits are represented as follows:

$$Q = \left\langle \begin{array}{c} 1/\sqrt{2}|1/\sqrt{2}|1/\sqrt{2} \\ 1/\sqrt{2}|1/\sqrt{2}|1/\sqrt{2} \end{array} \right\rangle \tag{4}$$

A Q-bit individual has the ability to represent a linear superposition of states. Thus, a Q-individual with three-Q-bits can represent the linear superposition of eight states, which is presented as follows:

$$\begin{aligned} Q = {} & (1/2\sqrt{2})\langle 000\rangle + (1/2\sqrt{2})\langle 001\rangle + (1/2\sqrt{2})\langle 010\rangle \\ & + (1/2\sqrt{2})\langle 011\rangle + (1/2\sqrt{2})\langle 100\rangle + (1/2\sqrt{2})\langle 101\rangle \\ & + (1/2\sqrt{2})\langle 110\rangle + (1/2\sqrt{2})\langle 111\rangle \end{aligned} \tag{5}$$

A Q-bit individual with three-Q-bit system would perform the operation on eight values, thus Q-bit individual with n-Q-bit will perform operations on 2^n values. As shown in Eq. (5), a single Q-bit individual is enough to represent eight states. Thus, Q-bit representation provides linear superposition of states in the search space probabilistically (Han and Kim 2002). Therefore, it has a better characteristic of population diversity than other representations that also enable us to exploit the global solution in the search space. Inspired by the concept of Q-bit representation in quantum computing, in this paper, we propose a Quantum-Inspired Evolutionary Fuzzy c-Means algorithm, which finds the best value of fuzziness factor and the number of clusters for the effective clustering of data from a large quantum search space.

3 Fuzzy c-Means (FCM) algorithm

Fuzzy c-Means algorithm is one of the well-known fuzzy clustering method developed by Dunn (1973) and improved by Bezdek (2013). It attempts to partition the set of n data points $X = [x_1, x_2, \ldots, x_n]$ into a collection of c fuzzy clusters $F = [F_1, F_2, \ldots, F_c]$ by minimizing the following objective function:

$$J_m(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij})^m \|x_i - v_j\|^2, m > 1 \tag{6}$$

where, μ_{ij} is the membership degree of data point x_j to the fuzzy cluster F_i, m is the fuzziness factor which controls the fuzziness of the resulting clusters. V=$[v_1, v_2, \ldots, v_c]$ is a vector of cluster centroids for the fuzzy clusters $F_1, F_2, \ldots, F_c$. The goal of the FCM algorithm is to iteratively improve a set of fuzzy cluster centroids until no further change in $J_m(U,V)$. The step-wise procedure of FCM algorithm is stated as follows:

Algorithm 1: **Fuzzy c-Means**

1: **Randomly** initialize the number of clusters (c), $c_{\min} \leq c \leq c_{\max}$ where $c_{\min} = 2$ and $c_{\max} = \sqrt{n}$, the cluster centroid $V = [v_1, v_2.., v_c]$ and the fuzziness factor m ($m > 1$) and set termination criteria $T = 0.001$.
2: **Compute** the cluster membership matrix $U_i = [\mu_{ij}]$ for $i = 1, 2, ..., c$.

$$\mu_{ij} = \frac{\| x_j - v_i \|^{\frac{-2}{m-1}}}{\sum_{k=1}^{c} \| x_j - v_k \|^{\frac{-2}{m-1}}} \tag{7}$$

3: **Check** the fuzzy partition matrix U_i obtained in Eq. (7) satisfy the condition stated below:

$$\sum_{i=1}^{c} \mu_{ij} = 1 \tag{8}$$

4: **Compute** the fuzzy cluster centroid v_i for $i = 1, 2, .., c$.

$$v_j = \frac{\sum_{i=1}^{n}[(\mu_{ij})^m]x_j}{\sum_{i=1}^{n}(\mu_{ij})^m} \tag{9}$$

5: **If** improvement in $J_m(U,V)$ is less than T, then stop; otherwise go to step 2.

The performance of the Fuzzy c-Means algorithm is sensitive to the selection of c and m which is done randomly. Due to this, FCM will not be able to guarantee of getting close to the optimal results and likely to fall into local optima (Flores-Sintas et al. 1999).

4 Fuzzy cluster validity index

In fuzzy clustering, finding the optimal number of clusters is an important problem. The cluster validity index is widely adapted to find the optimal number of clusters when the number of clusters is not known in advance. Consequently, the cluster validity is also used to validate the quality of produced partitions such that whether the produced partitions accurately present the structure of the dataset or not. The literature suggests that the most important factors in the validity indexes are intra-cluster compactness, inter-cluster separation, and inter-cluster overlap. The intra-cluster compactness refers to the density of data points present in the cluster. The higher compactness indicates the good partition. In contrast, the inter-cluster separation indicates that how far apart the clusters are located from each other. The higher value of this term indicates the larger separation between the clusters and better fuzzy partitions. The other important factor inter-cluster overlap indicates the degree of overlap between the fuzzy clusters. In the literature, a wide variety of cluster validity indexes exist for fuzzy clustering. In this section, we review some of the cluster validity indexes available in the literature.

4.1 Partition coefficient (PC) and partition entropy (PE)

Bezdek proposed two cluster validity indexes for fuzzy clustering, the partition coefficient (PC) (Bezdek 1974) and the partition entropy (PE) (Bezdek 1973), which is defined as:

$$V_{PC} = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^2 \tag{10}$$

$$V_{PE} = -\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}\log_a \mu_{ij} \tag{11}$$

Both the indexes achieve the optimal number of clusters (c) by maximizing the V_{PC} and minimizing the V_{PE} with respect to $c = [c_{\min}, \ldots, c_{\max}]$.

4.2 V_{CWB} validity index

Rezaee et al. (1998) proposed a validity index which is defined as follows:

$$V_{CWB} = \alpha \frac{\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}(x_j - v_i)^2}{cn||\sigma(X)||} + \frac{D_{\max}}{D_{\min}}\sum_{k=1}^{c}\left(\sum_{z=1}^{c}||v_k - v_z||\right)^{-1} \tag{12}$$

In this equation, $\sigma(X)$ denote the variance of pattern set X, and $D_{\min}$ and $D_{\max}$ are the minimum and maximum distances between the cluster centroids, respectively. The minimum value of V_{CWB} leads to an optimal value of c.

4.3 V_{OS} validity index

Kim et al. (2004) proposed a validity index which is defined as follows:

$$v_{os} = \frac{\frac{2}{c(c-1)} \sum_{p=1}^{c-1} \sum_{q=p+1}^{c} \times \left[\sum_{\mu} \sum_{j=1}^{n} \delta(x_j, \mu : \tilde{F_p}, \tilde{F_q}) \omega(x_j) \right]}{1 - \underbrace{\min}_{p \neq q} \left[\underbrace{\max}_{x \in X} \min(\mu_{\tilde{F_p}}(x), \mu_{\tilde{F_q}}(x)) \right]} \tag{13}$$

A minimum value of this index indicates an optimal value of c with the partition in which clusters overlap to a lesser degree and are more separated from each other.

4.4 VI_{DSO} validity index

Bharill and Tiwari (2014) proposed a validity index that validates the fitness of produced fuzzy partitions by three measures, i.e. intra-cluster compactness, inter-cluster separation, and inter-cluster overlap. The VI_{DSO} validity index is defined as follows:

$$VI_{DSO}(c, U) = \frac{Disp^N(c, U) + Overlap^N(c, U)}{Sep^N(c, U)} \tag{14}$$

$$Disp^N(c, U) = \frac{Disp(c, U)}{Disp_{\max}} \tag{15}$$

$$Sep^N(c, U) = \frac{Sep(c, U)}{Sep_{\max}} \tag{16}$$

$$Overlap^N(c, U) = \frac{Overlap(c, U)}{Overlap_{\max}} \tag{17}$$

where, $Disp(c, U)$ denotes the dispersion of all the data points for c number of clusters, $Sep(c, U)$ represents the separation among all pairs of c fuzzy clusters and $Overlap(c, U)$ denote the total overlap among all pairs of c fuzzy clusters. $Disp_{\max}$, $Sep_{\max}$, and $Overlap_{\max}$ represent the maximum value of dispersion, separation, and overlap. $Disp^N(c, U)$, $Sep^N(c, U)$, and $Overlap^N(c, U)$ show the normalized value of dispersion, separation, and overlap for c number of clusters. The minimum value of $VI_{DSO}(c, U)$ leads to an optimal value of c which indicates the partition in which clusters are more compact, overlapped with a lesser degree, and are more separated from each other.

5 Proposed Quantum-Inspired Evolutionary Fuzzy c-Means algorithm (QIE–FCM)

Pal and Bezdek (1995), suggested that fuzziness factor (m) and the number of clusters (c), significantly influence the clustering performance of FCM. To find an appropriate value of m and c for the efficient execution of FCM, in this paper, we propose a Quantum-Inspired Evolutionary Fuzzy c-Means (QIE–FCM) algorithm. In QIE–FCM, the fuzziness factor (m) is represented in terms of quantum bits (Q-bit). In QIE–FCM, quantum gates are used, which update the quantum bits to generate the different value of m in each generation from a large search space. For each value of m, the optimal number of clusters (c) is determined which corresponds to the minimum value of the fitness function (Bharill and Tiwari 2014). The working of the proposed QIE–FCM algorithm is categorized into five major steps: initialization of fuzziness factor (m), transformation process, selection of the objective function, evaluation of a fitness function, and evolution of fuzziness factor.

5.1 Initialization of fuzziness factor

Inspired by the idea mentioned above of Han and Kim (2002), in the proposed algorithm, the quantum concept is utilized to find out the global optimal value of m from a large search space. To achieve this, the fuzziness factor (m) is represented in terms of quantum bits and a different value of m is evolved in each generation of the QIE–FCM. Let us represent the fuzziness factor (m) in terms of quantum bits for a generation (g) as M'_g, which is defined as follows:

$$M'_g = (Q^g) \tag{18}$$

where, a Q-bit individual Q^g is assumed to contain k quantum bits, which are represented as $Q^g = [\alpha_1^g \mid \alpha_2^g \mid \cdots \mid \alpha_k^g]$. The best value of fuzziness factor (m) is obtained from a large search space which is divided into 2^k subspaces. However, the proposed algorithm run on a classical computer; therefore, there is a need to convert the fuzziness factor from quantum bits M'_g to the real coded value m_g. This conversion can be done with the help of the transformation process (Lu et al. 2013).

5.2 Transformation process

The transformation process is used to obtain the real coded parameter m_g from the quantum value of fuzziness factor (M'_g). It starts by taking random R where R= $[r_1, r_2, \ldots, r_k]$, corresponding to $Q^g = [\alpha_1^g \mid \alpha_2^g \mid \cdots \mid \alpha_k^g]$. The value of r_i is selected with the help of random function which generates uniform number between 0 and 1. Then, the further mapping is done by using the binary matrix S^g where S^g = $[s_1^g \ldots s_i^g \ldots s_k^g]$. The value of matrix S^g is generated as follows:

$$if(r_i \leq (\alpha_i^g)^2) \quad \text{then}\, s_i^g = 1 \quad \text{else}\, s_i^g = 0. \tag{19}$$

Now, with the help of binary matrix S^g and the Gaussian random generator (grg) having parameters mean value $\bar{\mu}_i^g$ and variance $(\sigma_i^g)^2$, is represented as $grg(\bar{\mu}_i^g, \sigma_i^g)$, the real coded value of fuzziness factor (m_g) has been generated. The observation process shows the process of conversion of a single Q-bit. The step-wise procedure of the transformation process is discussed as follows:

Transformation process()

Input: Initialize $M_g^{'}$ and $link = 0$.
Output: m_g
Begin
for $i := 1$ to k **step** 1 **do**
  $Q^g = \alpha_i^g;\quad 0 \leq \alpha_i^g \leq 1$
  $r_i = rand$;
end for
for $i := 1$ to k **step** 1 **do**
  if $r_i \leq (\alpha_i^g)^2$
    $s_i^g = 1$;
  else
    $s_i^g = 0$;
  end if
end for
$link = bin2dec(S^g) + 1$
if $link \sim= 0$
  $i = link$;
  $m_g = grg(\bar{\mu}_i^g, \sigma_i^g)$;
end if
return
End

The transformation process can be understood with the help of an example. Let a quantum bit of length two-Q-bits are represented as $Q = \langle 0.345|.877\rangle$ and a random number matrix is generated using a random number R=[0.65 0.08]. Now using Eq. (19), the binary matrix is generated as S=[01]. Once the binary matrix is achieved, then the formula is used which convert a binary number to a decimal value (bin2dec(S)+1). This return a number between 1 to 4 and corresponding four Gaussian random values, for example grg(0.25, 0.03), grg(0.40, 0.03), grg(0.55, 0.03), grg(0.70, 0.03). As a binary value achieved here is S=[01] which return 2 as decimal value, therefore the real coded value from grg(0.40, 0.03) corresponding to quantum bit Q is selected.

5.3 Evaluation of local and global fitness function

Once the real coded value of fuzziness factor (m_g) is obtained through the transformation process, then for the different values of $c = [c_{\min},\ldots, c_{\max}]$, the FCM algorithm is executed. For each value of c, after several iterations of the FCM algorithm, the stable cluster centroids and the corresponding fuzzy partitions are obtained. The selection of best fuzzy partition from a set of obtained fuzzy partitions is a critical issue. In this paper, the VI_{DSO} index proposed by Bharill and Tiwari (2014) is used as the objective function to evaluate the fitness of obtained fuzzy partitions. The value of VI_{DSO} objective function on different values of c has different scales that need to be reconciled through a normalization process. Thus we normalized the VI_{DSO} objective function for each value of c, which is defined as follows:

$$VI_{DSO}^{sum}(c, U) = \sum_{c=c_{\min}}^{c_{\max}} VI_{DSO}(c, U) \tag{20}$$

$$VI_{DSO}^{Normalized}(c, U) = \frac{VI_{DSO}(c, U)}{VI_{DSO}^{sum}(c, U)} \tag{21}$$

To ensure the selection of best fuzzy partition in generation (g), the evaluation of local fitness function is done by finding the minimum value from the set of normalized values of $VI_{DSO}(c, U)$ objective function, which is denoted by $F_{Lbest}^g(m_g, c)$ and represented as follows:

$$F_{Lbest}^g(m_g, c) = \min_{c_{\min} \leq c \leq c_{\max}} [VI_{DSO}^{Normalized}(c, U)] \tag{22}$$

where $F_{Lbest}^g(m_g, c)$ is the minimum value of the fitness function which is determined corresponding to m_g by varying the value of clusters $c = [c_{\min}, \ldots, c_{\max}]$. Furthermore, to ensure the selection of the best value of fuzziness factor (m) from the large search space provided by the quantum concept, one more parameter is taken into account, which stores the global best value of the fitness function among all the generations denoted by $F_{Gbest}(m_{best}, c_{best})$. Let us assume, the initial value of $F_{Gbest}(m_{best}, c_{best})$ as ∞. The formulation of global best fitness function $F_{Gbest}(m_{best}, c_{best})$ is defined as follows:

$$F_{Gbest}(m_{best}, c_{best}) = \min(F_{Gbest}(m_{best}, c_{best}), F^{g}_{Lbest}(m_g, c)) \quad (23)$$

where m_{best} and c_{best} denote the best value of fuzziness factor (m) and the number of clusters (c) generated through an evolutionary process from a large quantum search space. As discussed above, to find the best value of m, the QIE–FCM algorithm is executed for several generations. To do so, there is a need for evolving the fuzziness factor (m) in each generation. Thus, the new value of m in each generation (g) is obtained by updating quantum bits of fuzziness factor (M'_g). To update the quantum bits of fuzziness factor from M'_g to M'_{g+1}, quantum rotation gates (Han and Kim 2004) are used which is described in the subsequent section.

5.4 Evolution of fuzziness factor

To evolve the different value of fuzziness factor in each generation, the quantum bits of fuzziness factor are updated using the quantum rotation gates (Han and Kim 2004). The quantum gates require proper angle to rotate the quantum bit, which is defined as follows:

$$U(\Delta\theta) = \begin{vmatrix} \cos(\Delta\theta) - \sin(\Delta\theta) \\ \sin(\Delta\theta)\cos(\Delta\theta) \end{vmatrix} \quad (24)$$

where $\Delta\theta$ is a rotation angle which is used to generate M'_{g+1} from M'_g. In the proposed algorithm, the length of a Q-bit individual $Q^g = [\alpha^g_1 \mid \alpha^g_2 \mid \cdot \mid \alpha^g_k]$ is considered as 2 (k=2). It means that m_g will be selected from 4 subspaces.

$$\begin{vmatrix} \alpha_i^{g+1} \\ \beta_i^{g+1} \end{vmatrix} = \begin{vmatrix} \cos(\Delta\theta) - \sin(\Delta\theta) \\ \sin(\Delta\theta)\cos(\Delta\theta) \end{vmatrix} * \begin{vmatrix} \alpha_i^{g} \\ \beta_i^{g} \end{vmatrix} \quad (25)$$

where rotation angle ($\Delta\theta$) is calculated on the basis of global and local fitness function $F_{Gbest}(m_{best}, c_{best})$ and $F^g_{Lbest}(m_g, c)$. The transformation process shows that each Q-bit α^g_i is associated with the binary value s^g_i, therefore a mapping is done between $F_{Gbest}(m_{best}, c_{best})$, $F^g_{Lbest}(m_g, c)$ and s^g_i to update an individual Q-bit. The values of $F_{Gbest}(m_{best}, c_{best})$ and $F^g_{Lbest}(m_g, c)$ is obtained corresponding to s^{global}_i and s^g_i where, s^{global}_i and s^g_i are the binary values corresponding to $F_{Gbest}(m_{best}, c_{best})$ and $F^g_{Lbest}(m_g, c)$, respectively. Table 1, summarizes the above-discussed parameters. If the value of $F^g_{Lbest}(m_g, c)$ obtained in the current generation is worse than the value of $F_{Gbest}(m_{best}, c_{best})$ obtained in the previous generation, and state of s^g_i is zero in the current generation and s^{global}_i is one, then decrementing the probability of α^g_i to zero may produce the worst result. Therefore, to update α^g_i, it is required that $\Delta\theta$ must be negative. On the contrary, if the value of $F^g_{Lbest}(m_g, c)$ obtained in the current generation is better than the value of $F_{Gbest}(m_{best}, c_{best})$ obtained in the previous generation, and state of s^g_i is one in the current iteration and s^{global}_i is zero, then increasing the probability of α^g_i to one may produce the worst result. Therefore, to update α^g_i, $\Delta\theta$ must be positive. In other cases, $\Delta\theta$ will remain zero. The value of $\Delta\theta$ must be selected in such a way so that it can take a minimum number of iterations to cover a maximum number of values of α^g_i in the range of (0, 1). Therefore, $\Delta\theta$ must be initialized between [$0.01 \times \pi$, $0.05 \times \pi$] (Han and Kim 2004).

Table 1 Parameters for Qubits updation

s^g_i	s^{global}_i	$F_{Gbest}(m_{best}, c_{best}) > F^g_{Lbest}(m_g, c)$	$\Delta\theta$
0	0	False	0
0	0	True	0
0	1	False	$-\ 0.03{*}\Pi$
0	1	True	0
1	0	False	0
1	0	True	$0.03{*}\Pi$
1	1	False	0
1	1	True	0

For preventing the quantum bit α^g_i from acquiring values 0 or 1, following constraint, is applied:

$$\alpha^g_i = \begin{cases} \sqrt{\varepsilon}, & \text{if} \quad \alpha^g_i < \sqrt{\varepsilon} \\ \alpha^g_i & \text{if} \quad \sqrt{\varepsilon} \le \alpha^g_i \le \sqrt{1-\varepsilon} \\ \sqrt{1-\varepsilon} & \text{if} \quad \alpha^g_i > \sqrt{1-\varepsilon} \end{cases} \quad (26)$$

where, the value of ϵ is assigned a very small (approximately approaching to zero) so that it can cover maximum values in the range of (0, 1). The QIE–FCM algorithm is executed for the g_{max} number of generations, which is set as the stopping criteria. The reason behind executing the QIE–FCM algorithm for g_{max} generations is that the best value of m lies within this interval can be easily found within the maximum number of generations. If the algorithm is executed for more than g_{max} generations, then it will generate the similar values of m which leads to increase the computational overhead. The step-wise procedure of the proposed QIE–FCM algorithm is described in Algorithm 1 as follows:

Algorithm 1. QIE-FCM algorithm for evolving the m and c

Input: Input the dataset $X = [x_1, x_2, \ldots, x_n]$ and parameter $F_{Gbest}(m_{best}, c_{best}) = \infty$.
Output: Global best fuzziness factor and the number of clusters i.e m_{best} and c_{best}.

```
Begin
Initialize the current generation number g as 1 and set the maximum number of generation g to g_max.
while g ≤ g_max do
   Initialize the fuzziness factor (m) for generation (g) in the form of quantum bits as M'_g=[α^g_1|α^g_2].
   Call transformation process(M'_g): To obtain the real coded value represented as m_g corresponding to the quantum value M'_g.
   Initialize parameter related to the FCM see Table 2.
   for c := c_min to c_max step 1 do
      Randomly initialize the cluster centroid v_i for i = 1,2,...,c and criteria function J_{m_g}(U,V : X,m_g,c) = ∞.
      repeat
         Compute the fuzzy partition matrix U_i = [u_ij] for i = 1,2,...,c using Eq. (7).
         Check the fuzzy partition matrix U_i obtained in Eq. (7) satisfy the condition stated in Eq. (8).
         Compute the fuzzy cluster centroid v_i for i = 1,2,...,c using Eq. (9).
         Compute the criteria function J_{m_g}(U,V : X,m_g,c) to evaluate the fitness of obtained fuzzy partition.
```

$$J_{m_g}(U,V:X,m_g,c) = \sum_{j=1}^{n}\sum_{i=1}^{c}(\mu_{ij})^{m_g}\|x_j - v_i\|^2, \quad 1 < m_g < \infty \tag{27}$$

```
      until (J_{m_g}(U,V : X,m_g,c) ≥ T)
   end for
   Compute the objective function VI_DSO(c,U) using Eq. (14) to evaluate the fitness of obtained partitions corresponding to m_g and all the values of c.
   Compute the summation of VI_DSO(C,U) objective function corresponding to all values of c using Eq. (20).
   Compute the normalized value of objective function VI_DSO(c,U) for all values of c = [c_min,..,c_max] using Eq. (21).
   Compute local best fitness i.e. F^g_Lbest(m_g,c) using Eq. (22) that determines the best fitness value in generation(g).
   Compute the global best fitness denoted by F_Gbest(m_best, c_best) using Eq. (23) that determines the best value of fuzziness factor and the number of clusters from the overall generations.
   Update the quantum bits (M'_g) by using Table 1 and Eq. (24).
   Update g = g+1.
end while
return Best value of fuzziness factor and the number of clusters i.e. m_best and c_best.
End
```

6 Experimental evaluation of QIE–FCM

The proposed QIE–FCM algorithm is implemented in MATLAB computing environment and executed on MATLAB version R2014a on Intel(R) Xeon(R) E5-1607 Workstation PC with 64 GB of memory and running on the Windows 7 Professional operating system with a processing speed of 3.0 GHz. In this study, three datasets from the famous UCI machine learning repository (Lichman 2015) are utilized to compare the performance and effectiveness of the QIE–FCM with those of two well-known fuzzy based clustering indexes (Rezaee et al. 1998; Kim et al. 2004) and other evolutionary fuzzy based clustering algorithms (Bandyopadhyay and Maulik 2001; Hung et al. 2013; Wang et al. 2014; Xiao et al. 2010; Silva Filho et al. 2015; Bandyopadhyay 2011). Experimentation is carried out on these datasets with the parameters as described in Table 2.

6.1 Datasets description

This section provides a description of Pima Indian Diabetes, Liver Disorder, and SPECTF Heart datasets used in the experiments are summarized as follows:

Pima Indian Diabetes (PID) dataset: It consists of 768 data samples that belong to only female patients suffering from diabetes. It has 8 numerical features per sample. The dataset is partitioned into two classes where the first class is labeled as “negative to diabetes” and involves 65.10% of the dataset (500 samples) while the second one is labeled as “positive to diabetes” and involves 34.89% of the dataset (268 samples).

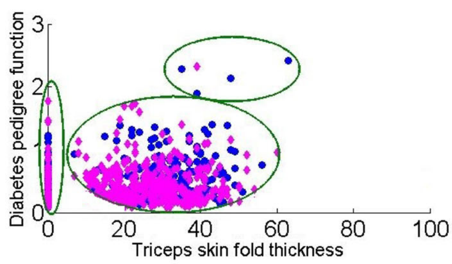
Liver Disorders dataset: This dataset is also named as BUPA Liver Disorders (BLD). It consists of 6 attributes in each data sample. The dataset consists of 345 data samples.

Table 2 Parameters used in QIE–FCM algorithm

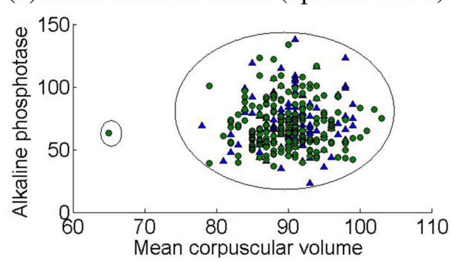
Parameters	Description	Values
T	Termination criteria	0.001
m	Fuzziness factor	[1.5, 2.5]
n	Number of input samples	Size of datasets
c	Number of clusters	$[c_{\min}, c_{\max}]$
$c_{\min}$	Minimum number of clusters	2
$c_{\max}$	Maximum number of clusters	$\sqrt{(n)}$
σ	Variance	0.6
$\Delta\Theta$	Rotation angle	$0.03*\Pi$
ϵ	limiting parameter	0.01
$g_{\max}$	maximum number of generations	100

These data samples are categorized into two classes where the first class involves 57.98% of the dataset (200 samples) while the second class involves 42.02% of the dataset (145 samples).

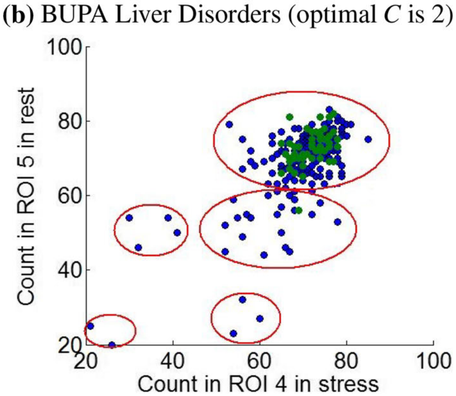
SPECTF Heart disease dataset: The dataset describes the diagnosis of cardiac Single Proton Emission Computed Tomography (SPECTF) images. It consists of 267 data samples and 44 numerical features per sample. The data samples are partitioned into two categories were first class labeled as “normal” and involved 79.40% of the dataset (212 sample) while the second class labeled as “abnormal” and involve 20.60% of the dataset (55 samples). Fig. 1a–c show scatters plot of the dataset with the optimal number of clusters (c) as per the distribution of data.



(a) Pima Indian Diabetes (optimal C is 3)



(b) BUPA Liver Disorders (optimal C is 2)



(c) SPECTF Heart (optimal C is 5)

Fig. 1 The scatter plots of Pima Indian Diabetes, BUPA Liver Disorder and SPECTF Heart Disease datasets with optimal number of clusters (C)

6.2 Results discussion

In this section, we present the experimental results to test the behavior of proposed QIE–FCM algorithm in comparison with well-known fuzzy based clustering indexes (Rezaee et al. 1998; Kim et al. 2004).

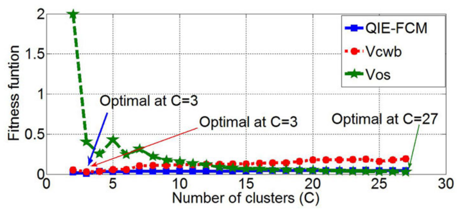
Table 3, present the optimal value of fitness function and the corresponding best value of fitness factor for all three datasets. For Pima Indian Diabetes dataset, QIE–FCM achieves the optimal fitness value 0.000031 at $m = 1.5$ which is much smaller than the optimal fitness value obtained by V_{CWB} and V_{OS} at $m = 1.5$ and $m = 2.5$, respectively. Similarly, for Liver Disorder dataset, the minimum fitness value achieved by QIE–FCM is 0.000037 at $m = 2.3$ which is remarkably lesser than the minimum fitness value, attain by V_{CWB} and V_{OS} at $m = 1.5$ and $m = 2.4$, respectively. In case of SPECTF heart dataset, the optimal fitness value achieved by QIE–FCM is 0.0000021 at $m = 2.4$ which is significantly much lesser than the minimum value of the fitness function obtained by V_{CWB} and V_{OS} at $m = 2.3$ and $m = 1.5$, respectively. The above-reported results justify the superiority of proposed QIE–FCM algorithm over classical validity indexes in terms of finding the global best value of fuzziness factor (m) and the corresponding value of the fitness function for all the datasets. The superiority of QIE–FCM over other indexes is mainly because the quantum concept is applied in fuzzy clustering for evolving fuzziness factor (m) in several generations. Through this, we get a larger search space, and thus we can assure the selection of the global optimal value of m and the fitness function by eliminating the problem of local convergence occurs in fuzzy clustering.

In Fig. 2 for the specific value of fuzziness factor ($m = 2$), we have reported the optimal number of clusters identified by our approach in comparison with two other indexes (Kim et al. 2004; Rezaee et al. 1998). According to the distribution of data shown in Fig. 1, the optimal number of clusters (c) for Pima Indian Diabetes, Liver Disorders, and SPECTF heart disease dataset is 3, 2, and 5, respectively. We can see from the results that, QIE–FCM achieves the minimum value of the fitness function for Pima Indian Diabetes, Bupa Liver Disorders, and SPECTF Heart Disease at $c = 3$, 2, and 5, respectively. As per the distribution of data shown in Fig. 1, the proposed approach correctly identifies the optimal number of clusters (c) for all the three datasets in comparison with V_{CWB} and V_{OS}. Hence, the reported results confirm the superiority of the proposed algorithm over classical indexes V_{CWB} and V_{OS} for all three datasets in terms of determining the optimal number of clusters (c).

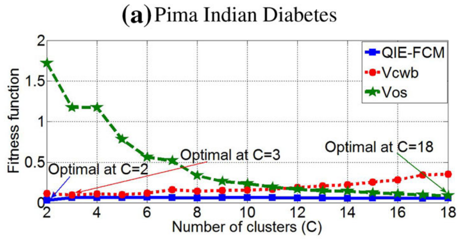
In Tables 4, 5 and 6, we have reported the number of iterations acquired by the QIE–FCM algorithm in determining the optimal number of clusters (c) corresponding to

Table 3 Comparison of optimal value of fitness function for three datasets

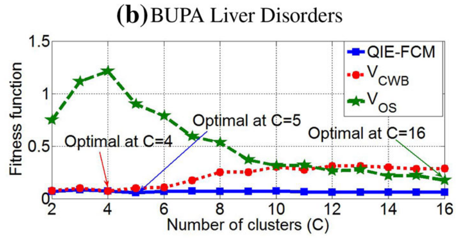
Datasets	QIE–FCM		V_{CWB} (Rezaee et al. 1998)		V_{OS} (Kim et al. 2004)	
	m	Fitness function	*m*	Fitness function	*m*	Fitness function
Pima Indian diabetes	1.5	0.000031	1.5	0.04130	2.5	0.06036
BUPA liver disorder	2.3	0.000037	1.5	0.05258	2.4	0.04233
SPECTF heart disease	2.4	0.0000021	2.3	0.00352	1.5	0.04475



(a) Pima Indian Diabetes



(b) BUPA Liver Disorders



(c) SPECTF Heart

Fig. 2 Comparison of number of clusters (c) evaluated by QIE–FCM algorithm, V_{CWB} and V_{OS} indexes at $m = 2$ for all the three datasets

the weighted exponent (m). The performance of the QIE–FCM algorithm is also compared with the V_{CWB} and V_{OS} indexes. It can be seen from results that for all the three datasets, the QIE–FCM algorithm always takes the least number of iterations with minimum standard deviation in comparison with V_{CWB} and V_{OS} indexes. This shows that QIE–FCM is computationally better than V_{CWB} and V_{OS} indexes.

6.3 Comparison with other methods

The proposed method is also compared with many fuzzy based evolutionary clustering algorithms (Bandyopadhyay and Maulik 2001; Hung et al. 2013; Wang et al. 2014; Xiao et al. 2010; Silva Filho et al. 2015; Bandyopadhyay 2011) in terms of the number of the clusters and value of the fitness function. It is observed from the results reported in Table 7 that the proposed approach was able to identify the exact number of clusters correctly for all the datasets. In contrast, the comparable approaches fail to find the exact number of clusters for these datasets. In addition to this, the optimal value of the fitness function achieved by the proposed approach is comparatively much lesser than the minimum value of the fitness function achieved by other compared approaches. It is important to highlight that proposed approach also identifies the appropriate value of

Table 4 Comparison of number of iterations for Pima Indian diabetes data

Datasets	Fuzziness factor	QIE–FCM		V_{CWB} (Rezaee et al. 1998)		V_{OS} (Kim et al. 2004)	
		Mean	Std dev	Mean	Std dev	Mean	Std dev
Pima Indian diabetes	1.51	160.03	0.8105	160.8	0.8234	161.57	0.8379
	1.99	155.73	0.9810	156.5	1.0043	157.26	1.0284
	2.25	164.00	1.0647	164.76	1.0854	194.34	1.5205
	2.30	156.23	0.9574	163.50	1.0295	164.26	1.0539
	2.32	168.61	1.0597	169.38	1.0844	170.15	1.1099
	2.37	197.11	1.2522	197.88	1.2715	198.65	1.2917
	2.40	166.15	1.0054	166.15	1.0054	195.34	1.5426
	2.45	162.38	0.9361	202.38	1.5659	209.00	1.7121
	2.48	148.42	0.8260	150.00	0.8700	164.76	0.8917
	2.50	159.42	0.8556	167.15	0.8979	168.30	0.9248

Table 5 Comparison of number of iterations for BUPA liver disorder data

Datasets	Fuzziness factor	QIE–FCM		V_{CWB} (Rezaee et al. 1998)		V_{OS} (Kim et al. 2004)	
		Mean	Std dev	Mean	Std dev	Mean	Std dev
BUPA liver disorders	1.51	108.82	0.5708	110.00	0.6045	111.17	0.6402
	1.99	121.23	0.6699	122.41	0.6851	123.88	0.7062
	2.25	118.58	0.8545	121.35	0.6970	122.52	0.7267
	2.30	130.29	0.9109	131.47	0.9401	132.64	0.9708
	2.32	112.29	0.6160	113.47	0.6494	114.64	0.6846
	2.37	116.64	0.6513	117.82	0.6798	119.00	0.7104
	2.40	118.35	0.6689	119.52	0.6951	128.94	0.9022
	2.45	121.35	0.7753	143.00	0.9896	144.17	1.0150
	2.48	119.05	0.4921	123.76	0.6986	140.58	0.7091
	2.50	119.76	0.5496	122.11	0.6079	123.29	0.6406

Table 6 Comparison of number of iterations for SPECTF heart disease data

Datasets	Fuzziness factor	QIE–FCM		V_{CWB} (Rezaee et al. 1998)		V_{OS} (Kim et al. 2004)	
		Mean	Std dev	Mean	Std dev	Mean	Std dev
Spectf heart	1.51	191.13	1.4365	193.13	1.483	197.13	1.5833
	1.99	134.73	0.7419	137.40	0.8050	139.40	0.8575
	2.25	106.73	0.4314	141.53	0.7869	163.20	1.1623
	2.30	105.86	0.5191	107.20	0.5497	159.00	1.1083
	2.32	99.730	0.5337	101.73	0.5978	162.40	1.1056
	2.37	108.46	0.8821	108.86	0.8960	133.93	0.8982
	2.40	88.860	0.4564	147.73	0.7774	150.40	0.8328
	2.45	90.530	0.2971	152.53	1.0990	154.53	1.1561
	2.48	92.600	0.3764	141.60	0.8762	145.60	0.9741
	2.50	104.53	0.4550	107.86	0.5410	133.60	0.8871

the fuzziness factor (m) corresponding to the optimal value of the fitness function for all the datasets. On the contrary, the authors of the compared approaches do not address the issue of identifying the appropriate value of fuzziness factor (m) for these datasets. Hence, the above-reported results in Table 7 quantify the effectiveness of the proposed approach over other fuzzy based evolutionary approaches.

6.4 Clinical significance of experimental results

Computer-aided detection and diagnosis (CAD) Systems are getting popular and increasingly being used by clinical experts for detection and interpretation of diseases. The CAD systems help in avoiding the possible errors made by the experts in the course of diagnosis and also the medical data can be examined in a shorter time (Petrick et al. 2013). Furthermore, it is also used by clinicians as a decision support in developing their diagnoses. Machine learning techniques are widely used to construct the CAD systems which help in decision making in the face of uncertainty. Fuzzy set theory based approaches play a vital role in dealing with the uncertainty in the medical data (Begum and Devi 2011). From the experimental results reported above, it can be observed that the proposed quantum-inspired evolutionary fuzzy c-means clustering approach efficiently handles the uncertainty and achieve better clustering results on medical data in comparison with other evolutionary fuzzy based clustering approaches. Thus, the clustering results achieved using the proposed approach in terms of the optimal number of clusters and the fuzziness factor helps in the formation of proper clusters which acts as an efficient decision support system to deal with medical data.

Table 7 Comparison with other methods from literature

Methods	Datasets					
	Pima Indian diabetes		Bupa liver disorder		SPECTF heart disease	
	Number of clusters (c)	Fitness function value	Number of clusters (c)	Fitness function value	Number of clusters (c)	Fitness function value
QIE–FCM	3	0.000031	2	0.000037	5	0.0000021
FCMVGA (Bandyopadhyay and Maulik 2001)	5	0.011029	3	0.040345	7	0.060573
QM–FCM (Hung et al. 2013)	16	0.00458	5	0.03072	9	0.00993
RQECA (Wang et al. 2014)	22	0.001174	15	0.003932	11	0.000119
KMQGA (Xiao et al. 2010)	2	15.78227	2	18.3299	2	10.496
KMVGA (Xiao et al. 2010)	2	11.97661	2	16.90276	2	7.162
FCM–PSO (Silva Filho et al. 2015)	2	78.3385	2	138.9046	2	199.6642
FCM–IDPSO (Silva Filho et al. 2015)	2	78.306	2	138.6978	2	199.6818
FCM–IDPSO2 (Silva Filho et al. 2015)	2	78.3057	2	138.0453	2	199.6642
HPSOFCM (Silva Filho et al. 2015)	2	88.8669	2	187.9076	2	215.4946
CPSFC (Silva Filho et al. 2015)	2	6.3192	2	58.2317	2	29.4008
GA clustering (Bandyopadhyay 2011)	2	2367.98	2	856.0932	2	15679.08

7 Conclusion

In this paper, we proposed a novel Quantum-Inspired Evolutionary Fuzzy c-Means algorithm for clustering of data. In this algorithm, the concept of quantum computing with Q-bit representation is employed to represent the fuzziness factor (m). The fuzziness factor (m) is evolved in each generation using five major steps (initialization of fuzziness factor, transformation process, selection of the objective function, evaluation of a fitness function, and Evolution of fuzziness factor). After several generations of evolution, we get a large search space that leads to the global optima and enables us to find the best value of m and c which results in the optimal fuzzy partition. The datasets used for experimentation are taken from the UCI Repository which is used to compare the performance of QIE–FCM with classical validity indexes V_{CWB} and V_{OS}. The QIE–FCM algorithm achieves the better value of the fitness function for Pima Indian Diabetes, BUPA liver disorders, and SPECTF Heart disease datasets are 0.000031, 0.000037, and 0.0000021, respectively. For these three datasets, the QIE–FCM finds the best value of the fuzziness factor which is 1.5, 2.3, and 2.4, respectively, and their exact number of clusters (c) are 3, 2, and 5 respectively. These results are much better than V_{CWB} and V_{OS} index.

In this paper, the reliability of QIE–FCM algorithm over other fuzzy based clustering indexes is also verified which show that the identified number of clusters (c) do not change with the change in values of m. In addition to this, the performance of the QIE–FCM algorithm is also compared to evolutionary fuzzy based clustering algorithms, and it is found that the proposed algorithm outperforms over these evolutionary approaches in terms of the number of clusters and the fitness function. From the above results, we conclude that our proposed QIE–FCM approach shows the competent or better results when compared with state-of-the-art methods.

In the aspect, concerning the limitation of this research, it is important to notice that the proposed approach is applied to the datasets is very small in size. In the practical scenario, a large volume of data with high dimensionality is generated nowadays. There are other areas like feature extraction, and feature selection will help in extracting the important features from the datasets and eliminate the redundant features making the QIE–FCM more efficient in clustering. However, it is part of our future work as discussed below.

The proposed work also opens future research directions in multiple topics, which include (but not limited to): (a) design of the scalable quantum-fuzzy clustering algorithm implemented on the Hadoop infrastructure for effective clustering of large data (b) To explore a large search space, the quantum concept can also be utilized for evolving the initial cluster centroid in the fuzzy clustering

(c) design of a hybrid quantum-genetic fuzzy clustering approach for exploring a large search space by evolving the initial cluster centroid using genetic and fuzziness factor using the quantum concept. Moreover, in future, the research can be carried out for the screening of features that can reduce the computational burden of the algorithm and result in efficient clustering of data. The proposed model is applied only to numeric data; it would be interesting to see its behavior when it is applied to different types of data available such as images, signals, text, and speech, etc.

References

Bandyopadhyay S (2011) Genetic algorithms for clustering and fuzzy clustering. Wiley Interdiscip Rev Data Min Knowl Discov 1(6):524–531

Bandyopadhyay S, Maulik U (2001) Nonparametric genetic clustering: comparison of validity indexes. IEEE Transact Syst Man Cybern Part C (Appl Rev) 31(1):120–125

Baskir MB, Turksen IB (2013) Enhanced fuzzy clustering algorithm and cluster validity index for human perception. Expert Syst Appl 40(3):929–937

Begum SA, Devi OM (2011) Fuzzy algorithms for pattern recognition in medical diagnosis. Assam Univ J Sci Technol 7(2):1–12

Bezdek J (1998) Pattern recognition in handbook of fuzzy computation. IOP Publishing Ltd., Boston

Bezdek JC (1973) Cluster validity with fuzzy sets. Taylor and Francis, London

Bezdek JC (1974) Numerical taxonomy with fuzzy sets. J Math Biol 1(1):57–71

Bezdek JC (2013) Pattern recognition with fuzzy objective function algorithms. Springer, Berlin

Bezdek JC, Ehrlich R, Full W (1984) Fcm: the Fuzzy c-Means clustering algorithm. Comput Geosci 10(2–3):191–203

Bharill N, Tiwari A (2014) Enhanced cluster validity index for the evaluation of optimal number of clusters for Fuzzy c-Means algorithm. In: IEEE international conference proceedings of the 2014 on fuzzy systems (FUZZ-IEEE), IEEE, pp 1526–1533

Dunn JC (1973) A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. Taylor and Francis, London

Erilli NA, Yolcu U, Egrioglu E, Aladag CH, Oner Y (2011) Determining the most proper number of cluster in fuzzy clustering by using artificial neural networks. Expert Syst Appl 38(3):2248–2252

Flores-Sintas A, Cadenas J, Martin F (1999) Membership functions in the Fuzzy c-Means algorithm. Fuzzy Sets Syst 101(1):49–58

Fukuyama Y, Sugeno M (1989) A new method of choosing the number of clusters for the Fuzzy c-Means method. In: Proceedings 5th fuzzy system symposium, vol 247, pp 247–250

Han KH, Kim JH (2002) Quantum-inspired evolutionary algorithm for a class of combinatorial optimization. IEEE Trans Evol Comput 6(6):580–593

Han KH, Kim JH (2004) Quantum-inspired evolutionary algorithms with a new termination criterion, h/sub/spl epsi//gate, and two-phase scheme. IEEE Trans Evol Comput 8(2):156–169

Hey T (1999) Quantum computing: an introduction. Comput Control Eng J 10(3):105–112

Höppner F (1999) Fuzzy cluster analysis: methods for classification, data analysis and image recognition. Wiley, New York

Huang M, Xia Z, Wang H, Zeng Q, Wang Q (2012) The range of the value for the fuzzifier of the Fuzzy c-Means algorithm. Pattern Recogn Lett 33(16):2280–2284

Hung CC, Casper E, Kuo BC, Liu W, Yu X, Jung E, Yang M (2013) A quantum-modeled Fuzzy c-Means clustering algorithm for remotely sensed multi-band image segmentation. In: Proceedings of the 2013 IEEE international geoscience and remote sensing symposium (IGARSS), IEEE, pp 2501–2504

Izakian H, Abraham A (2011) Fuzzy c-Means and fuzzy swarm for fuzzy clustering problem. Expert Syst Appl 38(3):1835–1838

Jain AK (2010) Data clustering: 50 years beyond k-means. Pattern Recogn Lett 31(8):651–666

Kaufman L, Rousseeuw PJ (2009) Finding groups in data: an introduction to cluster analysis. Wiley, New York

Kim DW, Lee KH, Lee D (2004) On cluster validity index for estimation of the optimal number of fuzzy clusters. Pattern Recogn 37(10):2009–2025

Li L, Liu X, Xu M (2007) A novel fuzzy clustering based on particle swarm optimization. In: Proceedings of the first IEEE international symposium on information technologies and applications in education, 2007 (ISITAE'07). IEEE, pp 88–90

Lichman M (2015) UCI machine learning repository, 2013. URL http://archive ics uci edu/ml 8

Lu TC, Yu GR, Juang JC (2013) Quantum-based algorithm for optimizing artificial neural networks. IEEE Transact Neural Netw Learn Syst 24(8):1266–1278

Ozkan I, Turksen I (2004) Entropy assessment for type-2 fuzziness. In: Proceedings of the 2004 IEEE international conference on fuzzy systems, 2004, vol 2. IEEE, pp 1111–1115

Pal NR, Bezdek JC (1995) On cluster validity for the fuzzy c-means model. IEEE Trans Fuzzy Syst 3(3):370–379

Petrick N, Sahiner B, Armato SG, Bert A, Correale L, Delsanto S, Freedman MT, Fryd D, Gur D, Hadjiiski L, Huo Z (2013) Evaluation of computeraided detection and diagnosis systems. Med Phys 40(8):1–16

Rezaee MR, Lelieveldt BP, Reiber JH (1998) A new cluster validity index for the Fuzzy c-Means. Pattern Recogn Lett 19(3):237–246

Silva Filho TM, Pimentel BA, Souza RM, Oliveira AL (2015) Hybrid methods for fuzzy clustering based on Fuzzy c-Means and improved particle swarm optimization. Expert Syst Appl 42(17):6315–6328

Stetco A, Zeng XJ, Keane J (2015) Fuzzy c-Means++: Fuzzy c-Means with effective seeding initialization. Expert Syst Appl 42(21):7541–7548

Wang H, Zhu W, Liu J, Li L, Yin Z (2014) Multidistribution center location based on real-parameter quantum evolutionary clustering algorithm. Math Probl Eng 2014:1–7

Webb AR (2003) Statistical pattern recognition. Wiley, New York

Wu KL (2012) Analysis of parameter selections for Fuzzy c-Means. Pattern Recogn 45(1):407–415

Xiao J, Yan Y, Zhang J, Tang Y (2010) A quantum-inspired genetic algorithm for k-means clustering. Expert Syst Appl 37(7):4966–4973

Xu R, Wunsch D (2005) Survey of clustering algorithms. IEEE Trans Neural Netw 16(3):645–678

Yu J, Cheng Q, Huang H (2004) Analysis of the weighting exponent in the fcm. IEEE Transact Syst Man Cybern Part B (Cybern) 34(1):634–639

Zhou K, Fu C, Yang S (2014) Fuzziness parameter selection in Fuzzy c-Means: the perspective of cluster validation. Sci China Inf Sci 57(11):1–8