Article

Rough IPFCM Clustering Algorithm and Its Application on Smart Phones with Euclidean Distance

Chih-Ming Chen [1], Sheng-Chieh Chang [2], Chen-Chia Chuang [3] and Jin-Tsong Jeng [2,*]

[1] School of Business, Fuzhou Institute of Technology, Fujian 350011, China; jmchen9684@gmail.com
[2] Department of Computer Science and Information Engineering, National Formosa University, Yunlin 632, Taiwan; qooqoo331@hotmail.com
[3] Department of Electrical Engineering, National Ilan University, Yilan 260, Taiwan; ccchuang@niu.edu.tw
[*] Correspondence: tsong@nfu.edu.tw

Abstract: New interval clustering technology for symbolic data analysis (SDA) on smart phones is shown to be beneficial for mobile computing devices for smart data analysis in this paper. A new interval clustering method that combined the rough set with interval possibilistic fuzzy C-means (IPFCM) algorithm under Euclidean distance is proposed and implemented on smart phones. Symbolic clustering algorithms (SCAs) have been widely used for pattern recognition, data mining, artificial intelligence, etc. In general, the SCA is unsupervised classification that is divided into groups according to symbolic data sets. However, the traditional interval fuzzy C-means (IFCM) clustering method still has noisy and data overlapping problems associated with these symbolic interval data. Hence, a new rough set with the interval possibilistic fuzzy C-means (RIPFCM) clustering algorithm with Euclidean distance was proposed to address the symbolic interval data (SID). That is, the proposed method can perform better than the traditional IFCM clustering algorithm for SID clustering in noisy environments and with data overlapping problems. The new RIPFCM algorithm under the Euclidean distance method was proposed to deal with SID on new applications in smart phones. Consequently, this method shows the expansion of the smart phone's computing power and its future application in new SDA.

Keywords: rough set; symbolic data analysis; fuzzy clustering; interval possibilistic fuzzy C-means clustering algorithm; smart phone

1. Introduction

In general, smart phones have better computing ability and connectivity to other devices than tradition phones. Their computing ability is similar to a personal computer in their ability to easily install and remove apps. Moreover, it is easy to use the internet via Wi-Fi and establish wireless connections via blue tooth and mobile broadband [1]. On the other hand, smart phones are small in size and easy and convenient to carry. In recent years, some popular mobile operating systems use Google's Android systems, Apple's iOS and Microsoft's Windows mobile systems [2]. The Windows mobile system (WMS) are developed to streamline the operating system for mobile products. That is, the WMS is designed to be as close as possible to the desktop version of Windows operating systems. At the same time, the WMS can, in combination with hardware, form the Windows mobile platform [3]. The .NET Compact Framework (CF) has two components, namely Common Language Runtime and .NET CF class library. At the same time, the .NET CF also has support Visual Basic .NET and C# language. Hence, Microsoft's WMS is easy to connect to Microsoft's Azure under C#. On the other hand, Azure app service mobile apps (AASMA) represents the development of integrating a rich set of features into mobile apps. It can accelerate the development of mobile applications. At the same time, it is easy to add data storage in the cloud, allows enterprise or social authentication, enables push notifications and offline syncing, and can scale to millions of devices simultaneously [1]. This paper

uses C#, which is compatible on AASMA for program development [4]. Hence, in this paper, we use C# language for the proposed interval clustering method to develop Windows mobile software applications. The era of the "touchscreen generation economy" is coming, and professional data analysis can be done at any time at the touch of a finger. Therefore, the development of AI methods on mobile platforms is currently a hot research field [5,6]. Kumari et al. [7] proposed the evaluation of machine learning and web-based processes for damage score estimation of existing buildings. Işık et al. [8] proposed precision irrigation systems using sensor network technology integrated with iOS/Android applications. Işık et al. [9] proposed the application of iOS/Android–based assessment and monitoring systems for building inventory under seismic impact. Harirchian et al. [10] proposed a prototype for the machine learning–based earthquake hazard safety assessment of structures by using a smartphone app.

In general, new interval clustering technology using the symbolic data analysis (SDA) method on smart phones is beneficial to mobile computing devices for smart data analysis. However, few studies have been published on this topic [11]. Hence, combining the rough set with the interval possibilistic fuzzy C-means (RIPFCM) algorithm under Euclidean distance is proposed and is implemented on smart phones in this paper. Symbolic clustering algorithms have been widely used for pattern recognition, data mining, artificial intelligence, etc. Therefore, this paper proposes a new interval fuzzy C-means clustering method to address symbolic interval data and the expansion of smart phones' computing power for new symbolic data analysis. That is, SDA and the expansion of smart phones' computing power is an important research topic in mobile computing and data analysis. The symbolic interval clustering algorithm involves unsupervised interval clustering that is divided into groups according to symbolic data sets. That is, analysis of interval data under fuzzy analysis requires a Type II fuzzy set [12] or interval fuzzy C-means (IFCM) clustering method [13]. Carvalho [13] extended the fuzzy C-means (FCM) clustering method and proposed an IFCM clustering method for symbolic interval data (SID) to conduct SDA. Jeng et al. [14,15] proposed an interval competitive agglomeration clustering algorithm. In general, the IFCM clustering method was proposed to deal with SID [13]. However, outliers and noisy problems are still associated with these SID. Chuang et al. [16] proposed a Hausdorff distance measure based an interval fuzzy possibilistic C-means (IFPCM) clustering algorithm for clustering SID. Jeng et al. [17] proposed an interval possibilistic FCM (IPFCM) clustering method for clustering SID. Because traditional IFPCM and IPFCM methods still have data overlapping problems for these SID, a novel IFCM clustering method that is roughly based on an interval possibilistic FCM (RIPFCM) clustering algorithm with Euclidean distance was proposed. The proposed method can outperform the IFCM clustering algorithm for SID clustering in noisy environments and for data overlapping problems. That is, we combined the rough set with the IPFCM algorithm under Euclidean distance for SID. We also extended the application of smart phones using machine learning and fuzzy clustering for SID. Finally, the contribution of the proposed RIPFCM clustering algorithm used the derived fuzzy membership degree (FMD) and possibilistic membership degree (PMD) formulas to enhance the proposed clustering method in SID that has noisy and data overlapping problems. An additional contribution is the successful implementation of the proposed clustering method on smart phones for new applications in mobile computing and data analysis.

The organization of the rest of the paper is as follows. In Section 2, a rough interval possibilistic fuzzy C-means clustering algorithm is proposed and discussed. The experimental results are shown in Section 3. Finally, the conclusions are summarized in Section 4.

2. Proposed Rough Interval Possibilistic Fuzzy C-Means Clustering Algorithm with Euclidean Distance

In general, fuzzy clustering and FCM clustering have been widely used in pattern recognition, data mining, machine learning, artificial intelligence, etc. [18–20], with single-valued data. In order to improve performance, it can be developed from the perspective of

rough set and FCM clustering integration. Hence, Zhao and Zho [21] proposed combining FCM clustering with fuzzy rough feature selection to conduct single-valued analysis. A rough set–based generalized FCM clustering algorithm was also proposed to address single-valued data [22]. In general, interval-valued data appear differently on single-valued data, representing uncertainty in the observed values. That is, the representation of interval-valued data is better able to handle the uncertainty of the data than single-valued data. Hence, the extended FCM clustering method which can deal with interval-valued data, is helpful for generalizing the applications of FCM clustering. In order to extend the interval-valued data analysis, a RIPFCM clustering algorithm with Euclidean distance data that extended [22] to interval-valued data analysis is proposed in this paper. The proposed RIPFCM clustering algorithm with Euclidean distance uses the concepts of the crisp lower approximation with $\underline{A}(g_i)$ and fuzzy boundary $B(g_i)$, as shown in Figure 1. This is, if the datasets of x_k for $k = 1, \ldots, n$ belong to crisp lower approximation with $\underline{A}(g_i)$ for $i = 1, \ldots, c$, the FMD u_{ik} and PMD t_{ik} are equal one; otherwise, the datasets belong to fuzzy boundary, and the FMD u_{ik} and PMD t_{ik} are between zero and one. Hence, the proposed RIPFCM clustering algorithm with Euclidean distance combines the rough set and IPFCM clustering algorithm with Euclidean distance in this paper.
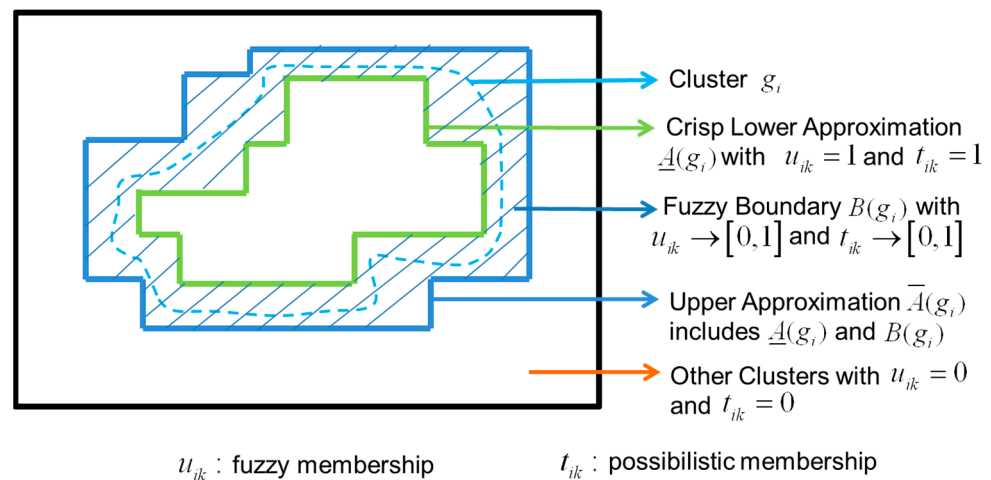

u_{ik} : fuzzy membership t_{ik} : possibilistic membership

Figure 1. The concept of the RIPFCM.

The new objective function J_{ripfcm} is based on a squared Euclidean distance between vectors of intervals; it includes the crisp lower approximation with $\underline{A}(g_i)$ and fuzzy boundary $B(g_i)$ given in Equation (1) and the constraints are given in Equation (2).

$$J_{ripfcm}(G, U, T, X) = \begin{cases} w \times A_1 + \widetilde{w} \times B_1 \ if \ \underline{A}(g_i) \neq \varnothing, \ B(g_i) \neq \varnothing \\ A_1, \qquad\qquad if \ \underline{A}(g_i) \neq \varnothing, \ B(g_i) = \varnothing \\ B_1, \qquad\qquad if \ \underline{A}(g_i) = \varnothing, \ B(g_i) \neq \varnothing \end{cases} \tag{1}$$

$$A_1 = \sum_{i=1}^{c} \sum_{k=x_k \in \underline{A}(g_i)}^{n} \left\{ q_1(u_{ik})^m + q_2(t_{ik})^\eta \right\} \phi_E \left(\vec{x}_k, \vec{g}_i \right) + \sum_{i=1}^{c} r_i \sum_{k=x_k \in \underline{A}(g_i)}^{n} (1 - t_{ik})^\eta$$

$$= \sum_{i=1}^{c} \sum_{k=x_k \in \underline{A}(g_i)}^{n} \left\{ q_1(u_{ik})^m + q_2(t_{ik})^\eta \right\} \sum_{j=1}^{p} \left(a_k^j - \alpha_i^j \right)^2 + \left(b_k^j - \beta_i^j \right)^2 + \sum_{i=1}^{c} r_i \sum_{k=x_k \in \underline{A}(g_i)}^{n} (1 - t_{ik})^\eta$$

$$= \sum_{i=1}^{c} \sum_{k=x_k \in \underline{A}(g_i)}^{n} (q_1 + q_2) \sum_{j=1}^{p} \left(a_k^j - \alpha_i^j \right)^2 + \left(b_k^j - \beta_i^j \right)^2$$

$$B_1 = \sum_{i=1}^{c} \sum_{k=x_k \in B(g_i)}^{n} \left\{ q_1(u_{ik})^m + q_2(t_{ik})^\eta \right\} \phi_E \left(\vec{x}_k, \vec{g}_i \right) + \sum_{i=1}^{c} r_i \sum_{k=x_k \in B(g_i)}^{n} (1 - t_{ik})^\eta$$

$$= \sum_{i=1}^{c} \sum_{k=x_k \in B(g_i)}^{n} \left\{ q_1 (u_{ik})^m + q_2 (t_{ik})^\eta \right\} \sum_{j=1}^{p} \left(a_k^j - \alpha_i^j \right)^2 + \left(b_k^j - \beta_i^j \right)^2 + \sum_{i=1}^{c} r_i \sum_{k=x_k \in B(g_i)}^{n} (1 - t_{ik})^\eta$$

$$\text{subject to } \sum_{i=1}^{c} u_{ik} = 1 . \tag{2}$$

In Equation (1), A_1 is the crisp lower approximation, B_1 is fuzzy boundary, w and $\widetilde{w}$ are the weight of crisp lower approximation and fuzzy boundary, $w + \widetilde{w} = 1$, $\varnothing$ is the empty set, u_{ik} is the FMD of pattern k in the ith cluster, $U = [u_{ik}]$ is a $c \times n$ matrix, $m \in [1, +\infty]$ is a fuzzy parameter that controls the FMD for each pattern k, t_{ik} is the PMD of pattern k in the ith cluster, $T = [t_{ik}]$ is a $c \times n$ matrix, $\eta \in [1, +\infty]$ is a parameter that controls the PMD for each pattern k, and r_i is the similarity of the possibilistic fuzzy C-means defined in [17] and given in Equation (3).

$$r_i = \frac{\sum\limits_{k=1}^{n} t_{ik}^\eta \sum\limits_{j=1}^{p} \left[\left(a_k^j - \alpha_i^j \right)^2 + \left(b_k^j - \beta_i^j \right)^2 \right]}{\sum\limits_{k=1}^{n} t_{ik}^\eta} . \tag{3}$$

In this paper, the Euclidean distance measure is used as an interval measure and defined as below:

$$\phi_E(\overrightarrow{x}_k, \overrightarrow{g}_i) = \sum_{j=1}^{p} \left[\left(a_k^j - \alpha_i^j \right)^2 + \left(b_k^j - \beta_i^j \right)^2 \right] \tag{4}$$

In Equation (4), $\phi_E(\overrightarrow{x}_k, \overrightarrow{g}_i)$ is the square of a suitable Euclidean distance that measures the dissimilarity between the vectors of the symbolic interval-valued data. Each pattern k is represented as a vector of intervals $\overrightarrow{x}_k = (x_k^1, \ldots, x_k^p)$ where $\overrightarrow{x}_k = \left[a_k^j, b_k^j \right]$ with $a_k^j \leq b_k^j$. The prototype $\overrightarrow{g}_i$ can be also represented as a vector of intervals $(g_i^1, \ldots, g_i^j, \ldots, g_i^p)$ where $\overrightarrow{g}_i^j = \left[\alpha_i^j, \beta_i^j \right]$ with $\alpha_i^j \leq \beta_i^j$ [13]. The core concept of rough set is as specified in [23,24], and the applications of rough set are as specified in [25,26]. In this paper, the proposed RIPFCM clustering algorithm combines the rough set, interval Euclidean distance measure, and IPFCM clustering algorithm.

The optimization problem of the proposed method can be addressed by minimizing the Lagrange multipliers method for the objective function J_{ripfcm}. The new unconstrained optimization problem is obtained as:

$$F_{ripfcm} = \sum_{i=1}^{c} \sum_{k=1}^{n} \left(q_1 u_{ik}^m + q_2 t_{ik}^\eta \right) \sum_{j=1}^{p} \left[\left(a_k^j - \alpha_i^j \right)^2 + \left(b_k^j - \beta_i^j \right)^2 \right] + \sum_{i=1}^{c} r_i \sum_{k=1}^{n} (1 - t_{ik})^\eta - \sum_{k=1}^{n} \lambda_k \left(\sum_{i=1}^{c} u_{ik} - 1 \right). \tag{5}$$

The FMD u_{ik} is considered for the minimum problem that $\frac{\partial F_{ripfcm}}{\partial u_{ik}}$ is equal to zero. Hence, to obtain an updated equation for the FMD u_{ik}, Equation (5) can be rewritten as follows:

$$m q_1 u_{ik}^{m-1} \sum_{j=1}^{p} \left[\left(a_k^j - \alpha_i^j \right)^2 + \left(b_k^j - \beta_i^j \right)^2 \right] - \lambda_k = 0$$
$$\text{for } i = 1, \ldots, c \text{ and } k = 1, \ldots, n \tag{6}$$

Therefore, the solution of Equation (6) for FMD u_{ik} can be rewritten as follows:

$$u_{ik} = \left(\frac{\lambda_k}{m q_1 \sum\limits_{j=1}^{p} \left[\left(a_k^j - \alpha_i^j \right)^2 + \left(b_k^j - \beta_i^j \right)^2 \right]} \right)^{\frac{1}{m-1}} , \tag{7}$$

In order to simplify FMD u_{ik}, Equation (7) can be substituted into Equation (2), as follows:

$$\sum_{s=1}^{c} \left(\frac{\lambda_k}{mq_1 \sum_{j=1}^{p}\left[\left(a_k^j - \alpha_s^j\right)^2 + \left(b_k^j - \beta_s^j\right)^2\right]} \right)^{\frac{1}{m-1}} = 1. \tag{8}$$

According to Equations (7) and (8), the updated equation for the FMD u_{ik} is obtained as follows:

$$u_{ik} = \left(\sum_{s=1}^{c} \left(\frac{\sum_{j=1}^{p}\left[\left(a_k^j - \alpha_i^j\right)^2 + \left(b_k^j - \beta_i^j\right)^2\right]}{\sum_{j=1}^{p}\left[\left(a_k^j - \alpha_s^j\right)^2 + \left(b_k^j - \beta_s^j\right)^2\right]} \right)^{1/(m-1)} \right)^{-1}, \tag{9}$$

The PMD t_{ik} is also considered for the minimum problem that $\frac{\partial F_{ripfcm}}{\partial t_{ik}}$ is equal to zero. Hence, to obtain an updated equation for the PMD t_{ik}, Equation (5) can be rewritten as follows:

$$\eta q_2 t_{ik}^{\eta-1} \sum_{j=1}^{p}\left[\left(a_k^j - \alpha_i^j\right)^2 + \left(b_k^j - \beta_i^j\right)^2\right] - \eta r_i (1 - t_{ik})^{\eta-1} = 0 \tag{10}$$

The solution of Equation (10) for PMD t_{ik} can be rewritten as

$$t_{ik} = \frac{1}{1 + \left(\frac{q_2 \sum_{j=1}^{p}\left[\left(a_k^j - \alpha_i^j\right)^2 + \left(b_k^j - \beta_i^j\right)^2\right]}{r_i} \right)^{\frac{1}{\eta-1}}}. \tag{11}$$

The prototype $\vec{g}_i = (g_i^1, \ldots, g_i^p)$ of the ith cluster, which minimizes the objective function J_{ripfcm}, has the bounds of the interval $\vec{g}_i^j = \left[\alpha_i^j, \beta_i^j\right] (j = 1, \ldots, p)$. Hence, the updated formula for the prototype $\vec{g}_i^j$ under the proposed method is as follows:

$$\vec{g}_{iripfcm}^j = \begin{cases} w \times C_1 + \widetilde{w} \times D_1 \ if \ \underline{A}(g_i) \neq \varnothing, \ B(g_i) \neq \varnothing \\ C_1, \qquad\quad if \ \underline{A}(g_i) \neq \varnothing, \ B(g_i) = \varnothing \\ D_1, \qquad\quad if \ \underline{A}(g_i) = \varnothing, \ B(g_i) \neq \varnothing \end{cases} \tag{12}$$

$$C_1 \Rightarrow \begin{cases} \frac{\partial A_1}{\partial \alpha_i^j} = 0 \Rightarrow -2(q_1 + q_2) \sum_{k=x_k\in\underline{A}(g_i)}^{n} \left(a_k^j - \alpha_i^j\right) = 0 \Rightarrow \alpha_i^j = \frac{1}{|\underline{A}(g_i)|} \sum_{k=x_k\in\underline{A}(g_i)}^{n} a_k^j \\ \frac{\partial A_1}{\partial \beta_i^j} = 0 \Rightarrow -2(q_1 + q_2) \sum_{k=x_k\in\underline{A}(g_i)}^{n} \left(b_k^j - \beta_i^j\right) = 0 \Rightarrow \beta_i^j = \frac{1}{|\underline{A}(g_i)|} \sum_{k=x_k\in\underline{A}(g_i)}^{n} b_k^j \end{cases}$$

$$D_1 \Rightarrow \begin{cases} \frac{\partial B_1}{\partial \alpha_i^j} = 0 \Rightarrow -2 \sum_{k=x_k\in B(g_i)}^{n} \left\{q_1(u_{ik})^m + q_2(t_{ik})^\eta\right\} \left(a_k^j - \alpha_i^j\right) = 0 \Rightarrow \alpha_i^j = D_1^{\alpha_i^j} \\ \frac{\partial B_1}{\partial \beta_i^j} = 0 \Rightarrow -2 \sum_{k=x_k\in B(g_i)}^{n} \left\{q_1(u_{ik})^m + q_2(t_{ik})^\eta\right\} \left(b_k^j - \beta_i^j\right) = 0 \Rightarrow \beta_i^j = D_1^{\beta_i^j} \end{cases}$$

where $D_1^{\alpha_i^j} = \frac{\sum_{k=x_k\in B(g_i)}^{n}\left\{q_1(u_{ik})^m + q_2(t_{ik})^\eta\right\}a_k^j}{\sum_{k=x_k\in B(g_i)}^{n}\left\{q_1(u_{ik})^m + q_2(t_{ik})^\eta\right\}}$ and $D_1^{\beta_i^j} = \frac{\sum_{k=x_k\in B(g_i)}^{n}\left\{q_1(u_{ik})^m + q_2(t_{ik})^\eta\right\}b_k^j}{\sum_{k=x_k\in B(g_i)}^{n}\left\{q_1(u_{ik})^m + q_2(t_{ik})^\eta\right\}}$.

Let $\mu_{ik} = (q_1 u_{ik} + q_2 t_{ik})$ and then let μ_{ik} and μ_{sk} be the highest and second highest memberships of x_k for $k = 1, \ldots, n$, respectively. If $(u_{ik} - u_{sk}) > \delta$, then $x_k \in \underline{A}(g_i)$, as well as $x_k \in \overline{A}(g_i)$; otherwise, $x_k \in \overline{A}(g_i)$, and $x_k \in \overline{A}(g_s)$, where δ is the weight of the rough degree. Algorithm 1 shows the procedure of the proposed RIPFCM clustering algorithm with Euclidean distance.

Algorithm 1 The procedure of the proposed RIPFCM clustering algorithm with Euclidean distance is shown in the following steps.

Step 1: Initialization for fix c between $2 \leq c < n$; fix m between $1 < m < \infty$; fix η between $1 < \eta < \infty$;
　　fix $\varepsilon > 0$; fix δ between $0 \leq \delta \leq (q_1 + q_2)$; fix w and $\widetilde{w}$, $w + \widetilde{w} = 1$;
　　set iteration counter to $l = 1$ and limit the number of iterations to L.
　　Initialize the FMD u_{ik}; initialize the PMD t_{ik}.
Step 2: Estimate r_i using Equation (3).
Step 3: Let $\mu_{ik} = (q_1 u_{ik} + q_2 t_{ik})$ for $i = 1, \ldots, c$ and $k = 1, \ldots, n$; then, let μ_{ik} and μ_{sk} be the highest and second highest
　　memberships of x_k for $k = 1, \ldots, n$.
　　If $(u_{ik} - u_{sk}) > \delta$, **then** $x_k \in \underline{A}(g_i)$, as well as $x_k \in \overline{A}(g_i)$;
　　otherwise, $x_k \in \overline{A}(g_i)$, and $x_k \in \overline{A}(g_s)$. That is,
　　if the datasets of x_k for $k = 1, \ldots, n$ belong to the crisp lower approximation with $\underline{A}(g_i)$ for $i = 1, \ldots, c$ **then** FMD u_{ik} and
　　　　PMD t_{ik} equal one;
　　otherwise,
they belong to the fuzzy boundary of FMD u_{ik} and PMD t_{ik} between zero and one.
Step 4: Update the prototypes $\overrightarrow{g}_i$ using Equation (12).
Step 5: Update the FMD u_{ik} and the PMD t_{ik} using Equations (9) and (10).
Step 6: **if** $\left(\left| J_{ripfcm}^{l-1} - J_{ripfcm}^{l} \right| \leq \varepsilon \text{ or } l > L \right)$, **then Stop**.
Otherwise, $l = l + 1$ and go to **Step 3**.

Remark 1. *The proposed method combines the rough set and IPFCM. The rough set can easily deal with data that includes overlapping problems in the data set. The IPFCM can easily deal with data that includes noise and outlier problems in the data set. Step 3 of Algorithm 1 involves judgement under the rough set to determine which formula to use. Step 4 and 5 of Algorithm 1 uses the derived formula. When data have no overlapping problems, such as in cases 1 and 3, the data set focuses on noise and outliers, and fewer data are needed to use the rough set. Hence, less improvement of the proposed method can be seen, as shown in Table 2. When data have overlapping problems, such as in cases 2 and 4, the data set focuses on overlapping problems, and more data are needed to use the rough set. Hence, a large improvement of the proposed method can be seen, as shown in Table 2.*

3. Experimental Results

　　Root mean squared error (*RMSE*) was used to calculate the true cluster center and corresponding cluster center, as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{c} (v_i - \widetilde{v}_i)^2}{c}}, \tag{13}$$

where v_i is the true cluster center and $\widetilde{v}_i$ is the corresponding cluster center. For further detail on the generation of SID, refer to [14,15,17]. Some preliminary results are published in the 22th International Symposium on Advanced Intelligent Systems [11]. The full derivation and experimental results are presented in this paper. The proposed experimental architecture is shown in Figure 2. There are four cases using SID in this paper. **Case 1** used three clusters in SID, and SID had no overlapping problems. In **Case 1,** SID had outliers and noise. Because different data sets were used, **Case 1** was divided into **Case 1a** and **Case 1b**. **Case 1a** considered the SID with outliers. Figure 3 shows the SID with outliers in **Case 1a**.

　　In order to build the interval data sets, they were herein defined as $([z1 - \gamma_1/2, z1 + \gamma_1/2], [z2 - \gamma_2/2, z2 + \gamma_2/2])$, from which γ_1 and γ_2 were drawn randomly from the intervals specified in [14,15], respectively. The name and version of the software developed in this paper is Mobile RIPFCM V1.0. The mobile platform displayed in Figure 4 has three subpages, namely the main windows page in (a), the initial results page in (b) and the final results page in (c). The main windows page shows the information on parameters, the initial results page shows the initial state of this case, and the final results page shows the results after the iteration of Algorithm 1. IC is represented as the iterated counter in Figure 4. OFV is represented as the value of the object function in Figure 4. The

design of the smart phones shown in Figures 6, 8, 13–16 is the same. The three colors of the center were red, green, and blue on the initial results page. The final results page also used the different colors to represent the results of each class of SID. The parameters of the RIPFCM clustering approach on the Windows mobile phone are $m = 2$, $\eta = 2$, $q_1 = 0.3$, $q_2 = 25$, and $\delta = 1.5$, the number of iterations is 300, and the number of clusters (nc) is 3 for **Case 1a**, as shown in Figure 4a. The initial state for the proposed mobile RIPFCM approach is shown in Figure 4b, and the final results are shown in Figure 4c for **Case 1a**.

Case 1b considered the SID with outliers and noise. Figure 5 shows the SID with outliers and noise for **Case 1b**.

The parameters of the RIPFCM approach on the Windows mobile phone are $m = 1.4$, $\eta = 1.7$, $q_1 = 0.3$, $q_2 = 18$, and $\delta = 1.6$, the number of iterations is 300, and the nc is 3 for **Case 1b** with noise and outliers, as shown in Figure 6a. The initial state for the proposed mobile RIPFCM approach is shown in Figure 6b, and the final results are shown in Figure 6c for **Case 1b** with noise and outliers.
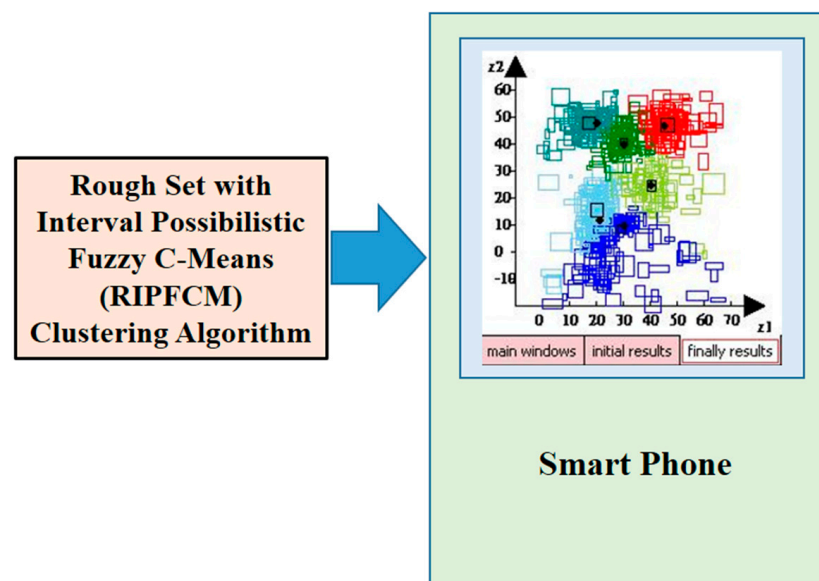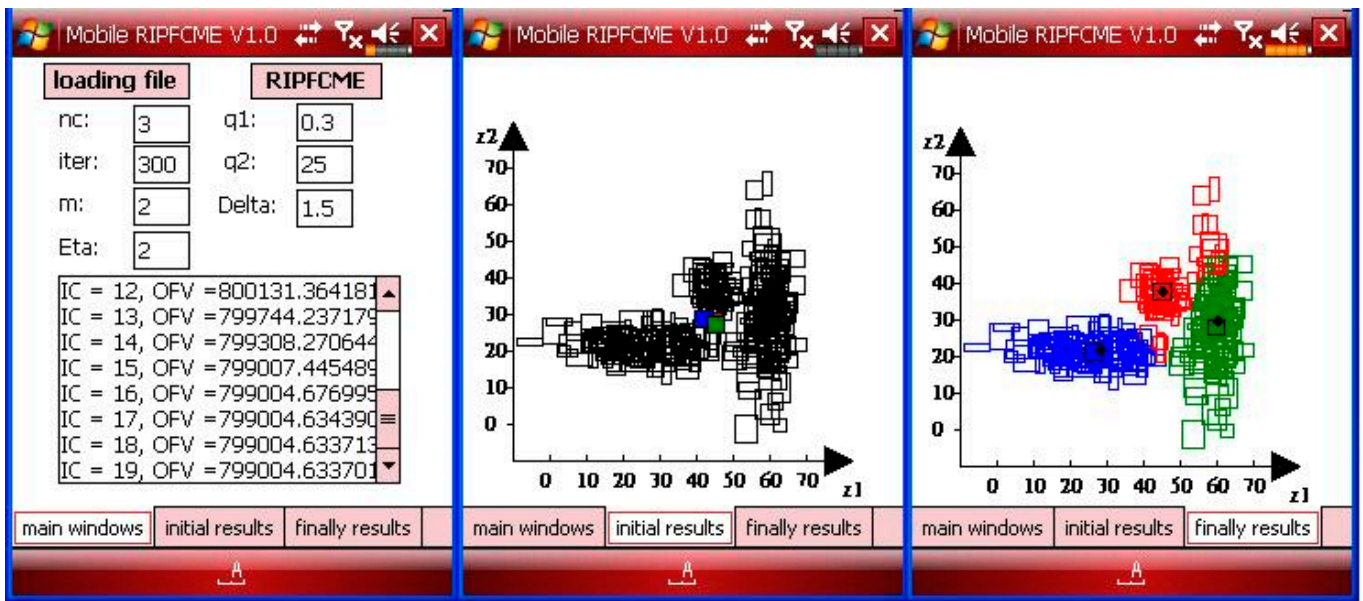

Figure 2. The proposed experimental architecture.
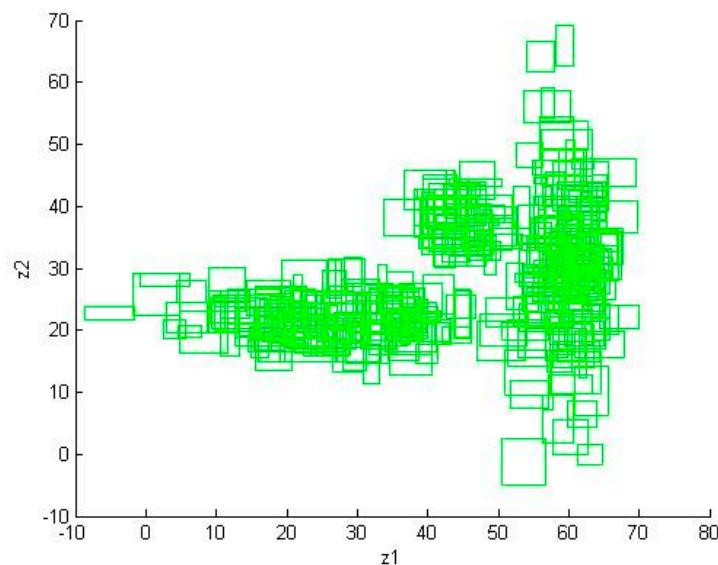
Figure 3. SID with outliers in **Case 1a**.

Figure 4. (**a**) The parameters of the main window, (**b**) the initial results window showing symbolic data with the initial center, and (**c**) the final results window showing the results with outliers for **Case 1a**.
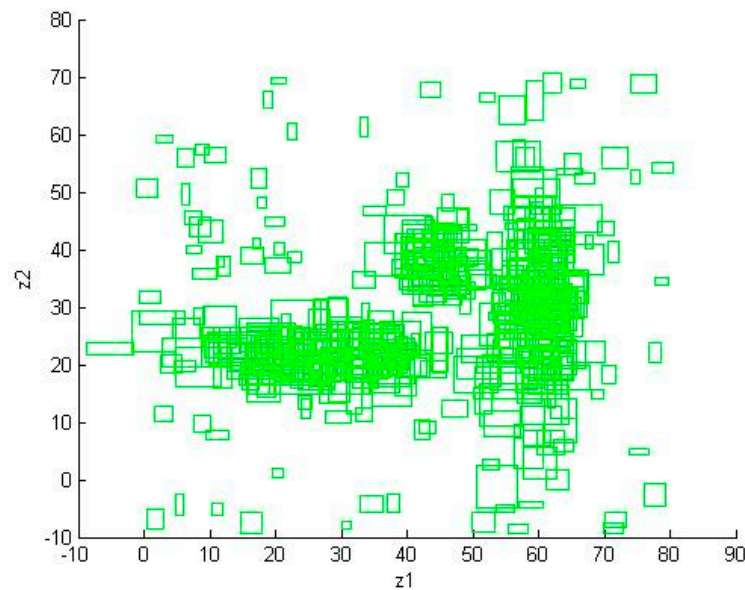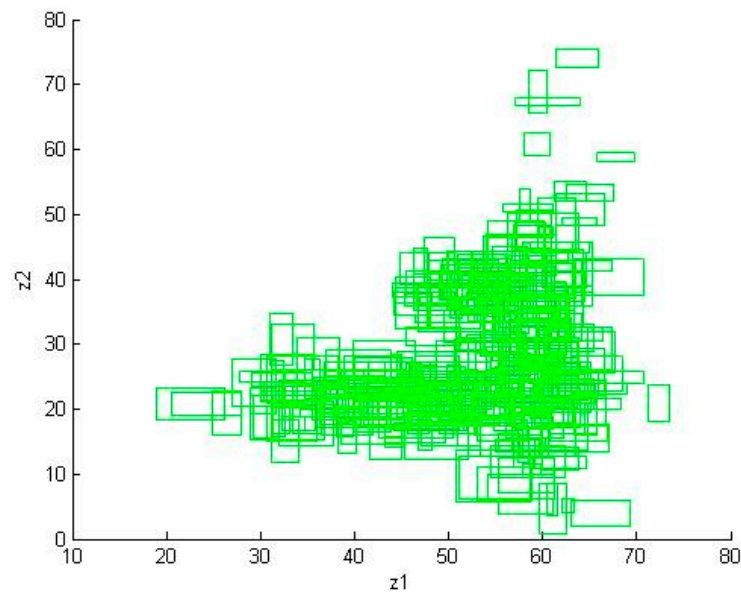

Figure 5. The SID with outliers and noise for **Case 1b**.

Case 2 used three clusters in SID, and the SID had overlapping problems. In **Case 2,** the SID had outliers and noise. **Case 2**, similar to **Case 1**, was divided into **Case 2a** and **Case 2b** in accordance with the different data sets. **Case 2a** represents SID with outliers. Figure 7 shows the SID with outliers for **Case 2a**.

The parameters of the RIPFCM approach on the Windows mobile phone are $m = 2$, $\eta = 1.5$, $q_1 = 0.5$, $q_2 = 25$, and $\delta = 0.9$, the number of iterations is 300, and the nc is 3 for **Case 2a**, as shown in Figure 8a. The initial state for the proposed mobile RIPFCM approach is shown n Figure 8b, and the final results are shown in Figure 8c for **Case 2a**.
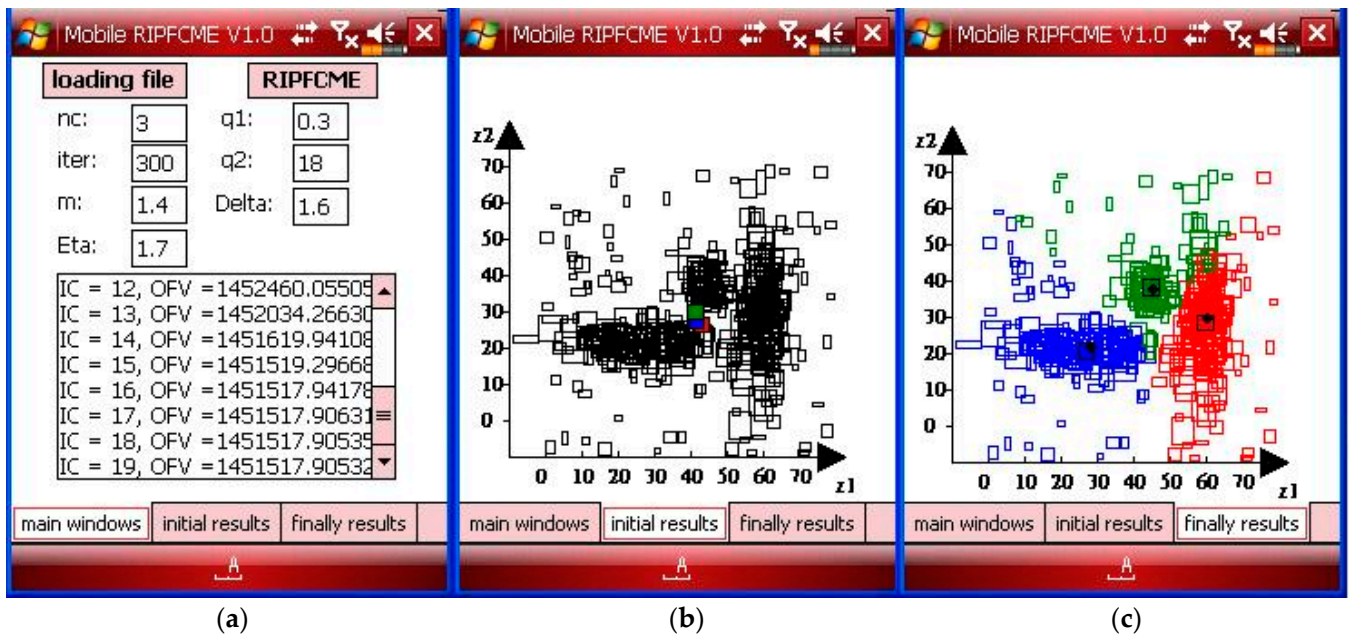
Figure 6. (**a**) The parameters of the main window, (**b**) the initial results window showing the symbolic data with the initial center, and (**c**) the final results window showing the results with outliers for **Case 1b**.

Figure 7. The SID with outliers for **Case 2a**.

Case 2b considered SID with outliers and noise, and the SID had overlapping problems. Figure 9 shows the SID with outliers and noise for **Case 2b**. In this case, a numerical example also illustrated the IFCM, IPFCM and RIPFCM under Euclidean distance.

We used Matlab to simulate this case. The number of iterations was 300. The results of IFCM with parameter $m = 2$ for **Case 2b** are shown on Figure 10. The results of IPFCM with parameters $m = 1.9$, $\eta = 2$, $q_1 = 1$, and $q_2 = 25$ for **Case 2b** are shown on Figure 11. The results of RIPFCM with parameters $m = 2$, $\eta = 1.3$, $q_1 = 0.3$, $q_2 = 22$, and $\delta = 3.5$ for **Case 2b** are shown on Figure 12. The RMSE results with IFCM, IPFCM, and RIPFCM under Euclidean distance are shown in Table 1.
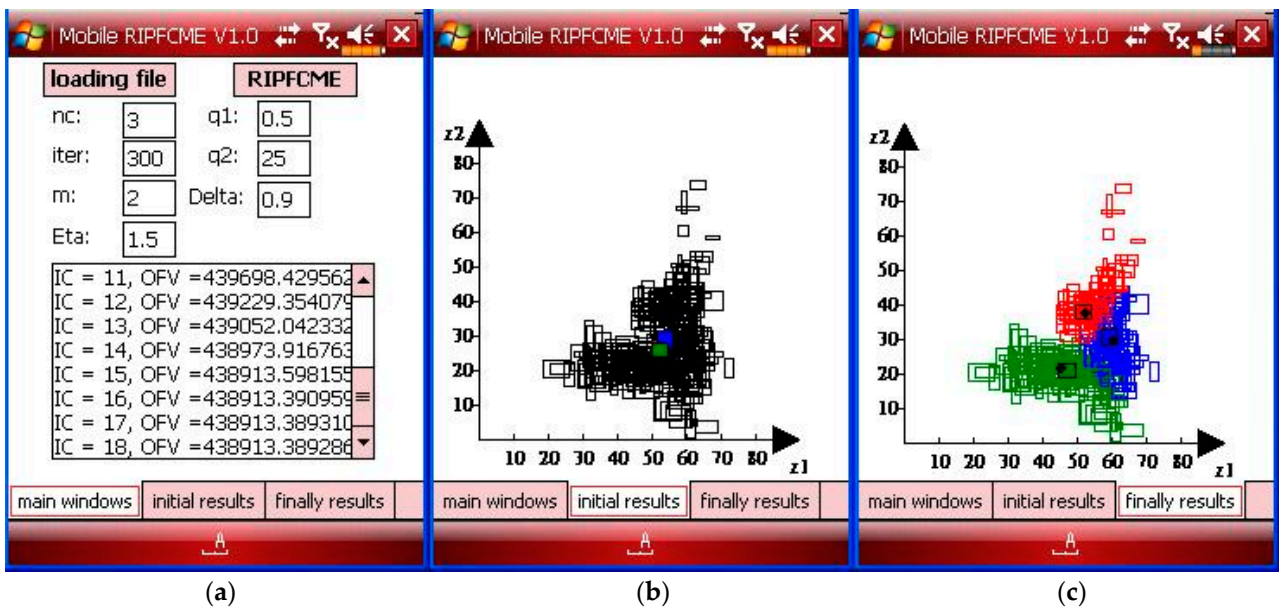
Figure 8. (**a**) The parameters of the main window, (**b**) the initial results window showing the symbolic data with the initial center, and (**c**) the final results window showing the results with outliers for **Case 2a**.
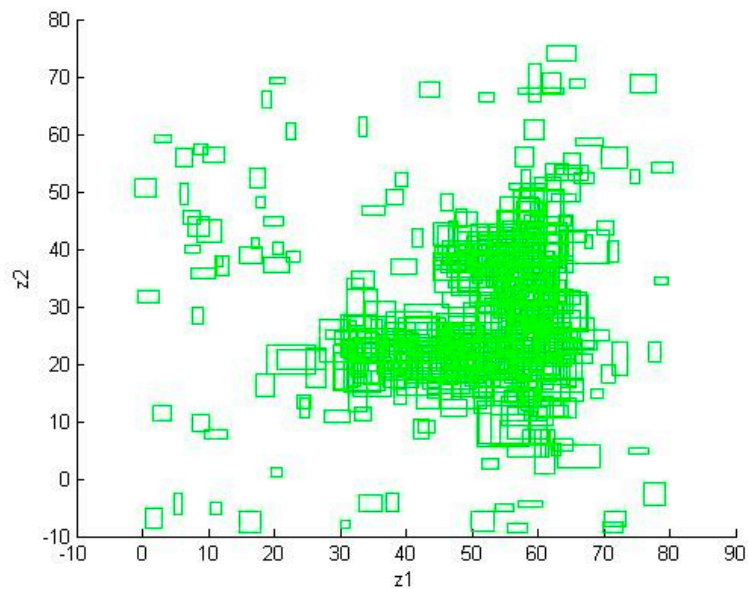

Figure 9. The SID with outliers and noise for **Case 2b**.

Table 1. Comparison of RMSE for IDS4 with different approaches.

	IFCM [9]	IPFCM [13]	RIPFCM
RMSE	10.2120	4.6490	1.1871

In this case, the parameters of the RIPFCM approach on the Windows mobile phone are $m = 2$, $\eta = 1.3$, $q_1 = 0.3$, $q_2 = 22$, and $\delta = 3.5$, the number of iterations is 300, and the nc is 3 for **Case 2b** with noise and outliers, as shown in Figure 13a. The initial state for the proposed mobile RIPFCM approach is shown in Figure 13b, and the final results are shown in Figure 13c for **Case 2b** with noise and outliers.
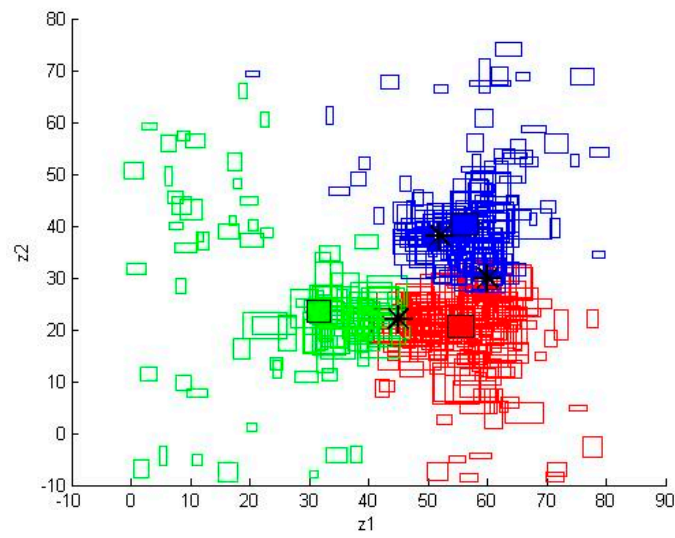
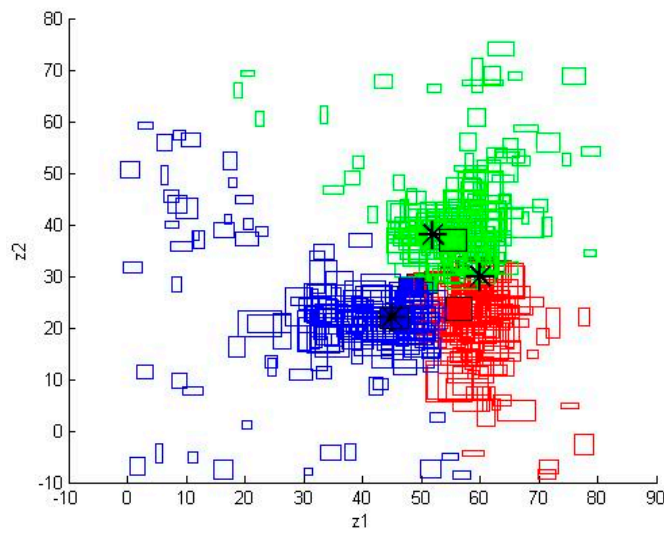**Figure 10.** The results of IFCM with parameter $m = 2$ for **Case 2b**.

Figure 11. The results of IPFCM with parameters $m = 1.9$, $\eta = 2$, $q_1 = 1$, and $q_2 = 25$ for **Case 2b**.
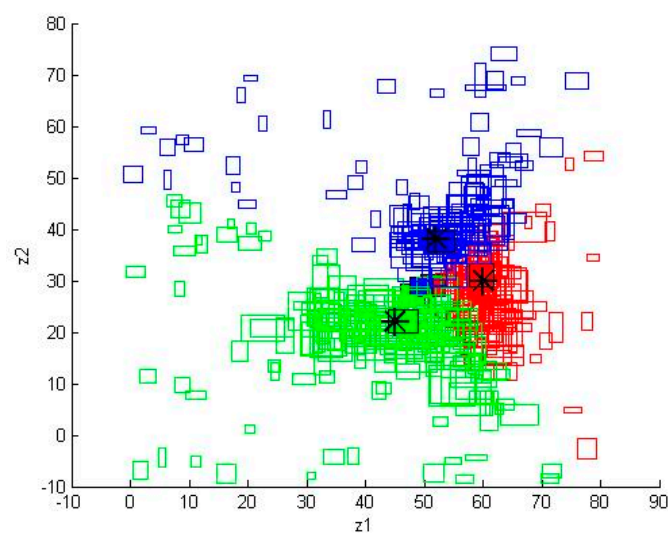

Figure 12. The results of RIPFCM with parameters $m = 2$, $\eta = 1.3$, $q_1 = 0.3$, $q_2 = 22$, and $\delta = 3.5$ for **Case 2b**.
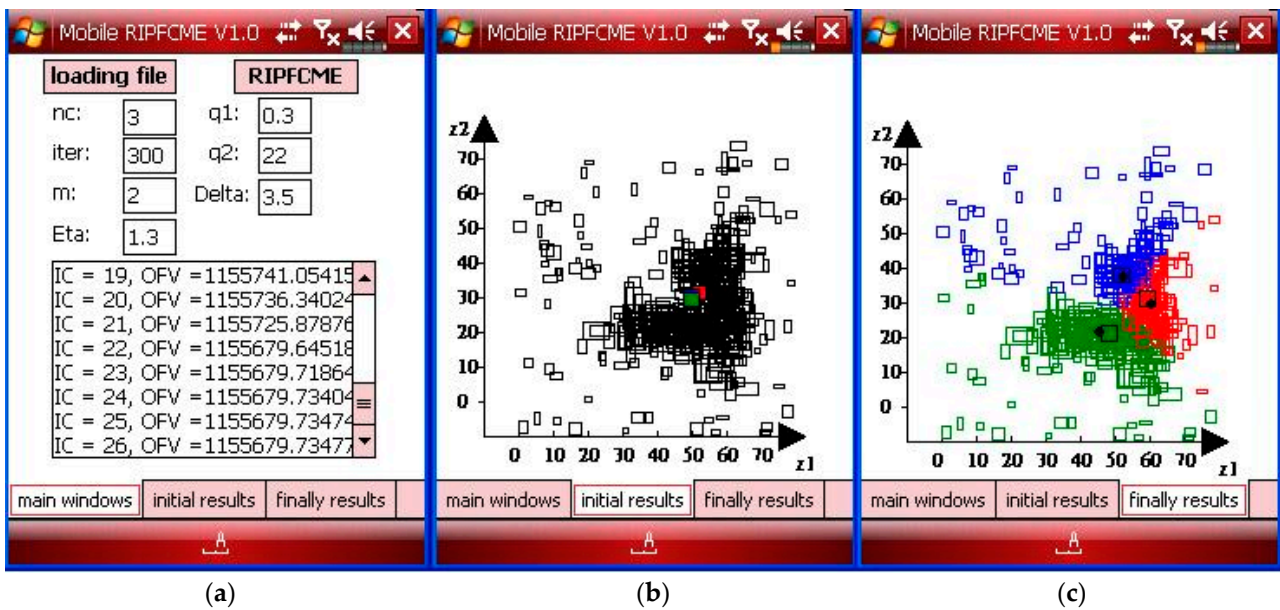
Figure 13. (**a**) The parameters of the main window, (**b**) the initial results window showing the symbolic data with the initial center, and (**c**) the final results window showing the results with outliers for **Case 2b**.

Case 3 used six clusters in SID, and the SID had no overlapping problem. In **Case 3**, the SID had outliers and noise. **Case 3a** considered SID with outliers. The parameters of the RIPFCM approach on the Windows mobile phone are $m = 1.2$, $\eta = 1.5$, $q_1 = 1$, $q_2 = 10$, and $\delta = 0.5$, the number of iterations is 300, and the nc is 6 for **Case 3a**, as shown in Figure 14a. The initial state for the proposed mobile RIPFCM approach is shown in Figure 14b, and the final results are shown in Figure 14c for **Case 3a**.
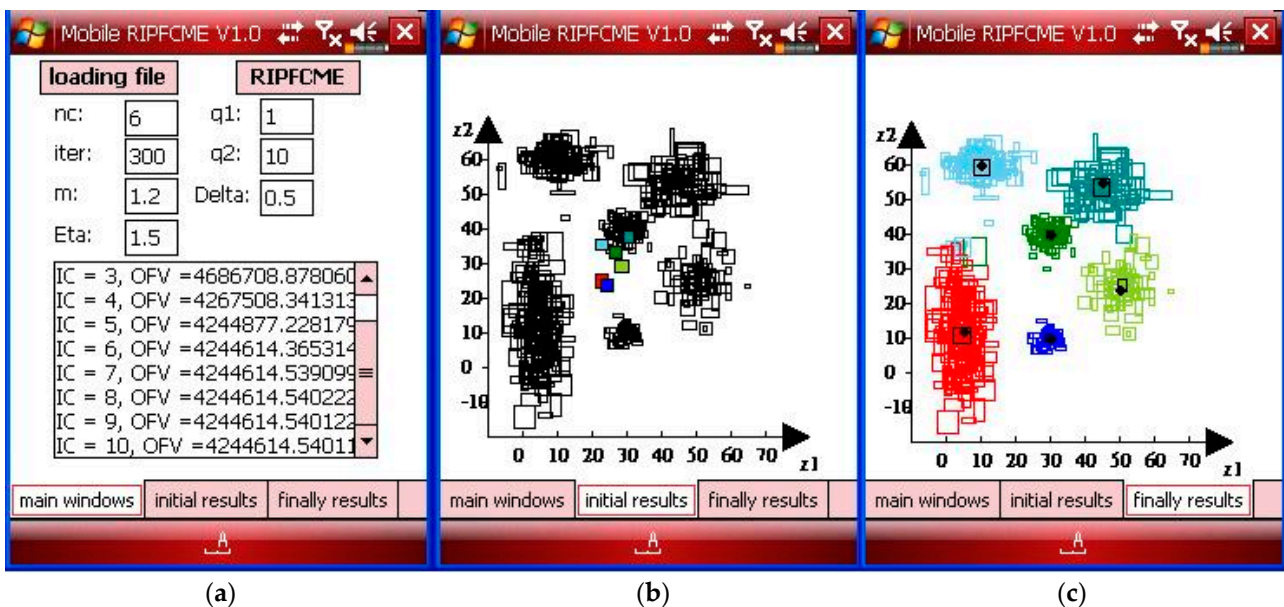

Figure 14. (**a**) The parameters of the main window, (**b**) the initial results window showing the symbolic data with the initial center, and (**c**) the final results window showing the results with outliers for **Case 3a**.

Case 3b considered SID with outliers and noise. The parameters of the RIPFCM approach on the Windows mobile phone are $m = 1.2$, $\eta = 2$, $q_1 = 1$, $q_2 = 8$, and $\delta = 1.5$,

the number of iterations is 300, and the nc is 6 for **Case 3b** with noise and outliers, as shown in Figure 15a. The initial state for the proposed mobile RIPFCM approach is shown in Figure 15b, and the final results are shown in Figure 15c for **Case 3b** with noise and outliers.

(a) (b) (c)

Figure 15. (**a**) The parameters of the main window, (**b**) the initial results window showing the symbolic data with the initial center, and (**c**) the final results window showing the results with outliers for **Case 3b**.

Case 4 used six clusters in SID, and the SID had overlapping problems. In **Case 4,** the SID consisted of a data set with outliers and noise for data overlapping problems. The parameters of the RIPFCM approach on the Windows mobile phone are $m = 1.1$, $\eta = 1.4$, $q_1 = 0.8$, $q_2 = 11$, and $\delta = 1.2$, the number of iterations is 300, and the nc is 6 for **Case 4** with noise and outliers, as shown in Figure 16a. The initial state for the proposed mobile RIPFCM approach is shown in Figure 16b, and the final results are shown in Figure 16c for **Case 4** with noise and outliers.
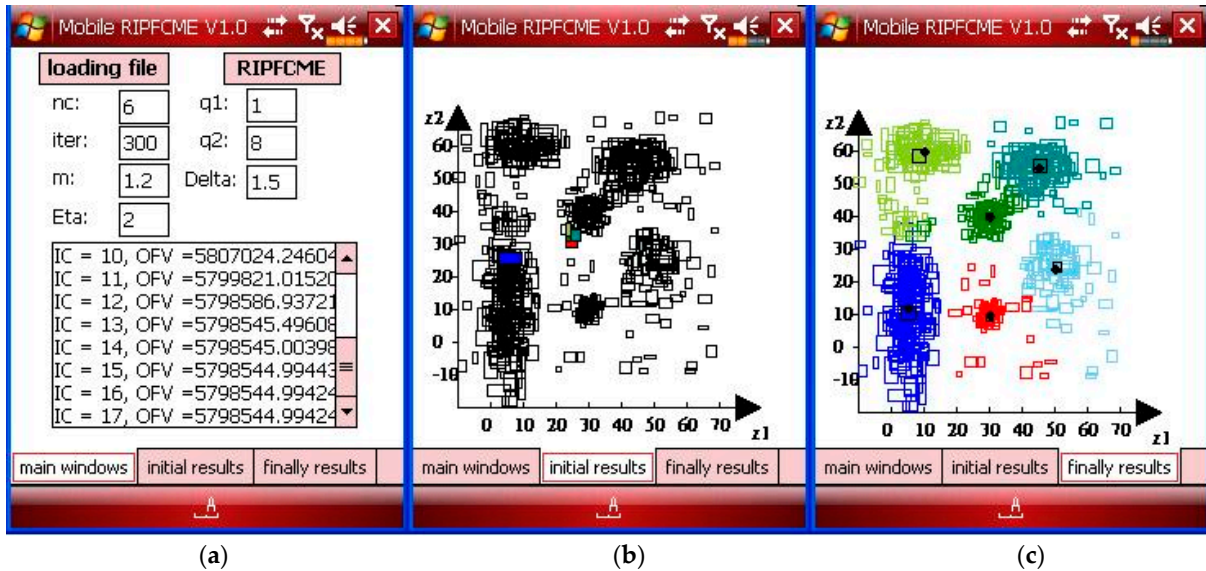
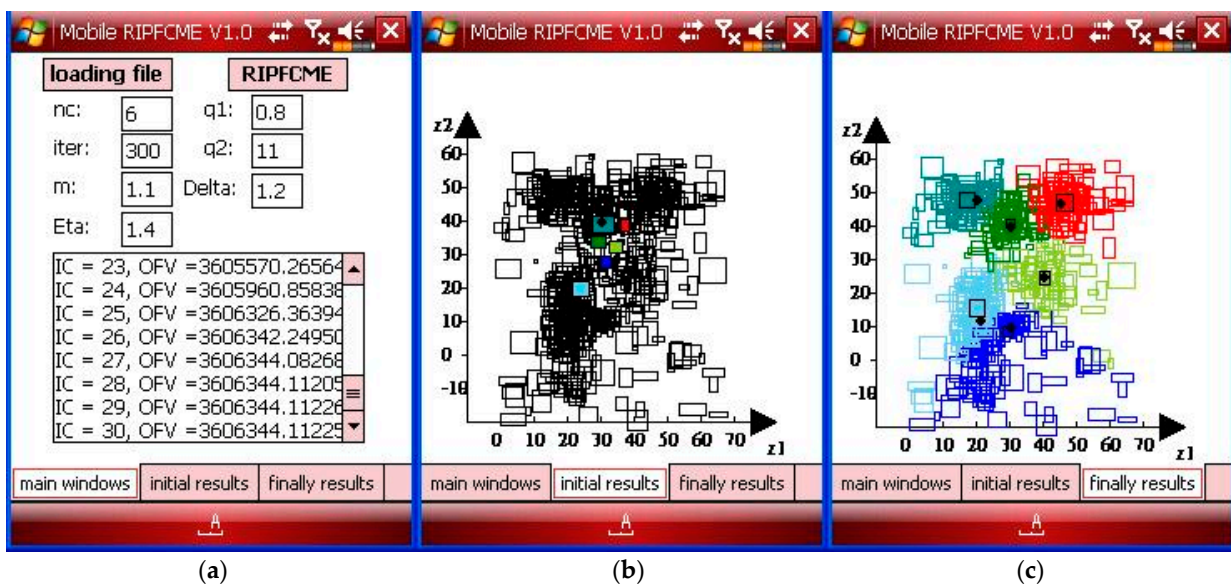

(a) (b) (c)

Figure 16. (**a**) The parameters of the main window, (**b**) the initial results window showing the symbolic data with the initial center, and (**c**) the final results window showing the results with outliers for **Case 4**.

Finally, Table 1 shows the comparisons of RMSE under four cases with different approaches. The validity of the results for the proposed method were verified.

From the simulation results, Tables 1 and 2 show that the proposed RIPFCM clustering algorithm has better performance than the IPFCM and IFCM clustering algorithm for the SID clustering with noise, outliers, and overlapping problems. At the same time, this also shows the expansion of smart phones' computing power and a possible future application in new symbolic data analysis. Hence, the proposed method was also implemented on a smart phone and extended the application of the smart phone in machine learning, fuzzy clustering, and non-single-valued data analysis. In order to establish the number of clusters to be used in each case, we used the IFCM clustering method for different numbers of clusters. Then, we used the most suitable results to test the proposed method in order to enhance the accuracy of the performance.

Table 2. Comparison of RMSE under four cases with different approaches.

RMSE	IFCM [9]	IPFCM [13]	Proposed Method
RMSE of Case 1a	2.9519	1.4371	1.0590
RMSE of Case 1b	4.0212	1.1363	0.7626
RMSE of Case 2a	6.3297	4.7523	1.5127
RMSE of Case 2b	10.2120	4.6490	1.1871
RMSE of Case 3a	1.6043	1.3731	1.0983
RMSE of Case 3b	—	1.0457	1.0457
RMSE of Case 4	—	1.8449	1.3827

— means poor convergence.

4. Conclusions

The new RIPFCM clustering algorithm with Euclidean distance is based on the integration of the rough set and the IPFCM clustering algorithm; it was applied to the analysis of SID as developed in this paper. The proposed new RIPFCM clustering algorithm not only can efficiently handle the overlapping partition problem but also shows better performance than the IPFCM clustering algorithm in the handing of noise problems for SID. In addition, we also combined the rough set to reduce the iteration number and increase the convergence of the IPFCM clustering algorithm. At the same time, the proposed new method also was implemented on a smart phone to address SID. This implementation showed the expansion of smart phones' computing power. Consequently, the proposed new method extended the application of smart phones in machine learning, fuzzy clustering, and non-single-valued data analysis and its future application in new SDA. Because the proposed method has more parameters, it has better accuracy performance with RMSE. However, its limitation is that it requires adjustment of the additional parameters, and thus the calculation speed of the proposed method will be slower.

Author Contributions: Conceptualization, S.-C.C. and J.-T.J.; data curation, C.-M.C. and C.-C.C.; methodology, C.-M.C., S.-C.C. and J.-T.J.; software, C.-M.C. and S.-C.C.; supervision, J.-T.J.; validation, C.-C.C.; visualization, C.-M.C. and S.-C.C.; writing—original draft, S.-C.C. and J.-T.J. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

Abbreviation	Meaning
AASMA	Azure app service mobile apps
CF	compact framework

FCM	fuzzy C-means
FMD	fuzzy membership degree
IFCM	interval fuzzy C-means
IFPCM	interval fuzzy possibilistic C-means
IPFCM	interval possibilistic FCM
PMD	possibilistic membership degree
RIPFCM	rough-based interval possibilistic FCM
RMSE	root mean squared error
SCA	symbolic clustering algorithm
SDA	symbolic data analysis
SID	symbolic interval data
WMS	Windows mobile system

References

1. Jeng, J.-T.; Chuang, C.-C.; Chang, S.-C. Interval fuzzy possibilistic c-means clustering algorithm on smart phone implement. In Proceedings of the 3rd Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), Sapporo, Japan, 9–12 September 2014.
2. Lee, J.; Kim, J.-W.; Lee, J. Mobile personal multi-access edge computing architecture composed of individual user devices. *Appl. Sci.* **2020**, *10*, 4643. [CrossRef]
3. Yao, P.; Durant, D. *Programming.NET Compact Framework 3.5*; Addison-Wesley Professional: Boston, MA, USA, 2009.
4. *ArcGIS Secure Mobile Implementation Patterns*; An Esri Software Security & Privacy Technical Paper; Esri: Redlands, CA, USA, 2021.
5. Pfeuffer, K.; Li, Y. Analysis and modeling of grid performance on touchscreen mobile devices. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; Volume 288, pp. 1–12.
6. Lin, Y.-C.; Wei, C.-C. Effects of touchscreen mobile devices and e-book systems for mobile users. In Proceedings of the 2017 International Conference on Applied System Innovation (ICASI), Sapporo, Japan, 13–17 May 2017.
7. Kumari, V.; Harirchian, E.; Lahmer, T.; Rasulzade, S. Evaluation of Machine Learning and Web-Based Process for Damage Score Estimation of Existing Buildings. *Buildings* **2022**, *12*, 578. [CrossRef]
8. Işık, M.F.; Sönmez, Y.; Yılmaz, C.; Özdemir, V.; Yılmaz, E.N. Precision Irrigation System (PIS) Using Sensor Network Technology Integrated with IOS/Android Application. *Appl. Sci.* **2017**, *7*, 891. [CrossRef]
9. Işık, M.F.; Işik, E.; Bülbül, M.A. Application of iOS/Android based assessment and monitoring system for building inventory under seismic impact. *J. Croat. Assoc. Civ. Eng.* **2018**, *70*, 1043–1056.
10. Harirchian, E.; Jadhav, K.; Kumari, V.; Lahmer, T. ML-EHSAPP: A prototype for machine learning-based earthquake hazard safety assessment of structures by using a smartphone app. *Eur. J. Environ. Civ. Eng.* **2021**, 1–21. [CrossRef]
11. Chang, S.-C.; Chuang, C.-C.; Jeng, J.-T. Rough IPFCM clustering algorithm and its application on smart phone with Euclidean distance. In Proceedings of the 22th International Symposium on Advanced Intelligent Systems, Cheonglu, Korea, 15–18 December 2021.
12. Mendel, J.M. *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, 2nd ed.; Springer: Cham, Switzerland, 2017.
13. Carvalho, F.D.A.D. Fuzzy c-means clustering methods for symbolic interval data. *Pattern Recognit. Lett.* **2007**, *28*, 423–437. [CrossRef]
14. Jeng, J.-T.; Chuang, C.-C.; Tao, C. Interval competitive agglomeration clustering algorithm. *Expert Syst. Appl.* **2010**, *37*, 6567–6578. [CrossRef]
15. Jeng, J.-T.; Chuang, C.-C.; Tseng, C.-C.; Juan, C.-J. Robust interval competitive agglomeration clustering algorithm with outliers. *Int. J. Fuzzy Syst.* **2010**, *12*, 227–236.
16. Chuang, C.-C.; Jeng, J.-T.; Chang, S.-C. Hausdorff distance measure based interval fuzzy possibilistic c-means clustering algorithm. *Int. J. Fuzzy Syst.* **2013**, *15*, 471–479.
17. Jeng, J.-T.; Chen, C.-M.; Chang, S.-C.; Chuang, C.-C. IPFCM clustering algorithm under Euclidean and Hausdorff distance measure for symbolic interval data. *Int. J. Fuzzy Syst.* **2019**, *21*, 2102–2119. [CrossRef]
18. Lu, Y.; Ma, T.; Yin, C.; Xie, X.; Tian, W.; Zhong, S. Implementation of the fuzzy c-means clustering algorithm in meteorological sata. *Int. J. Database Theory Appl.* **2013**, *6*, 1–18. [CrossRef]
19. Chui, K.T.; Lytras, M.D.; Vasant, P. Combined Generative Adversarial Network and Fuzzy C-Means Clustering for Multi-Class Voice Disorder Detection with an Imbalanced Dataset. *Appl. Sci.* **2020**, *10*, 4571. [CrossRef]
20. Yeom, C.-U.; Kwak, K.-C. Adaptive Neuro-Fuzzy Inference System Predictor with an Incremental Tree Structure Based on a Context-Based Fuzzy Clustering Approach. *Appl. Sci.* **2020**, *10*, 8495. [CrossRef]
21. Zhao, R.; Gu, L.; Zhu, X. Combining Fuzzy C-Means Clustering with Fuzzy Rough Feature Selection. *Appl. Sci.* **2019**, *9*, 679. [CrossRef]
22. Maji, P.K.; Pal, S.K. Rough Set Based Generalized Fuzzy CC -Means Algorithm and Quantitative Indices. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **2007**, *37*, 1529–1540. [CrossRef] [PubMed]
23. Pawlak, Z. *Rough Sets: Theoretical Aspects of Reasoning about Data*; Kluwer: Dordrecht, The Netherlands, 1991.
24. Pieta, P.; Szmuc, T. Applications of rough sets in big data analysis: An overview. *Int. J. Appl. Math. Comput. Sci.* **2021**, *31*, 659–683.

25.   Sahu, R.; Dash, S.R.; Das, S. Career selection of students using hybridized distance measure based on picture fuzzy set and rough set theory. *Decis. Making Appl. Manag. Eng.* **2021**, *4*, 104–126. [CrossRef]

26.   Sharma, H.K.; Kumari, K.; Kar, S. A rough set approach for forecasting models. *Decis. Making Appl. Manag. Eng.* **2020**, *3*, 1–21. [CrossRef]