Symbolic Clustering with Interval-Valued Data

Mika Sato-Ilic*

Faculty of Systems and Information Engineering, University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki 305-8573, Japan

Abstract

While many clustering techniques for interval-valued data have been proposed, there has been no proposal for a variable selection added fuzzy clustering method for high dimension low sample-size interval-valued data. This paper proposes such a novel fuzzy clustering method for interval-valued data with an adaptable variable selection. There are three reasons why the method is necessary: First, our target data in this study is high dimension low sample-size data. Due to the curse of dimensionality, we tend to obtain a poor classification result for this type of data. The main cause of this is noise occurring from irrelevant and redundant variables (dimensions). Therefore, we need to use an adaptable variable selection to reduce or summarize variables. Second, the merit of fuzzy clustering is to obtain the results with uncertain cluster boundaries, which is well adjusted with the uncertainty situation of classification to data. This gives a more robust result for the noise of data when compared with hard clustering while mathematically we can obtain a result with continuous values. Third, an adaptable representation of interval-valued data can be exploited to transform the original data into a more manageable data in order to avoid the curse of dimensionality. Numerical examples show a high performance for the proposed method.

©2011 Published by Elsevier B.V. Open access under CC BY-NC-ND license.

Keywords: symbolic data, interval-valued data, clustering, high dimension low sample-size data, subspace of variables

1. Introduction

Recently, analyses of high dimension low sample-size data in which the number of variables is much larger than the number of objects have gained tremendous interest from many researchers in various areas, such as genomics and other bioinformatics areas. Cluster analysis for such a data has the problem of poor classification caused by noise occurring from irrelevant and redundant variables (dimensions). In order to solve this problem, several cluster analyses have been proposed. [3], [4] However, the target data of these analyses is single-valued data, and there has been no proposal for interval-valued data. Interval-valued data is one type of symbolic data which has been discussed

*Corresponding author. Tel:+81-29-853-5006; fax:+81-29-853-5006
E-mail address: mika@risk.tsukuba.ac.jp
in symbolic data analysis [1], [2]. The essence of symbolic data analysis is to take the internal structure of data into analysis based on conceptual elements of the data structure. One merit of this analysis is to handle the general purpose representation of the data, which are represented by lists, intervals, distributions, etc. In particular, this paper discusses interval-valued representation. The merits of the representation of interval are that it is possible to reduce data size while retaining the knowledge within it and to represent the uncertainty of the data. Considering the noise of the high dimension low sample-size data, we exploit a feature of fuzzy clustering which has a robust solution.

Therefore, the purpose of this paper is to propose a novel clustering method which has two key features; flexibility of symbolic data representation and robustness of fuzzy clustering, in order to identify effective variables for the classification of the high dimension low sample-size symbolic data. This method derives three kinds of clusters; clusters which are given priori as external information of classification of objects in data, clusters of objects obtained by a clustering method, and clusters of variables obtained by a clustering method. In order to distinguish these kinds of clusters, we refer to these clusters in this paper as follows: a given cluster of objects as external information is referred to as a class, an obtained cluster of objects by a clustering method is referred to as a cluster, and obtained cluster of variables by a clustering method is referred to as a category.

This paper consists of the following: Section 2 proposes a variable selection criterion to reduce the number of variables based on external information of a classification of data. This method exploits an individual fuzzy classification based variable selection technique [7] and obtains a subspace spanned by variables which have enough explainable power for the classes of data given as external information of data. By using this criterion, we can select efficient variables for the classification. However, if the number of variables is very large when compared with the number of objects, the variable selection does not satisfactorily work to sufficiently reduce the number of variables while containing the internal knowledge of data. Therefore, in section 3, we propose a clustering method for interval variables in order to categorize the selected variables. In order to introduce the given class structure to this method, we transform from single-valued variables to interval-valued variables by expressing the classes as the intervals. Based on the categories of variables, section 4 explains how to obtain a manageable data form in which the number of objects is larger than the number of variables. In this method, we use interval representation of data in which uncertainty of variables for a category with respect to a fixed object is represented by an interval. For the clustering of the manageable formed data, we exploit fuzzy clustering. In fuzzy clustering, how to determine an adaptable number of clusters is important. So, in section 5, we describe a selection method for an adaptable number of clusters which has been proposed in our previous paper. [8] Section 6 describes a numerical example and section 7 contains conclusions.

2. Individual Classification Structure Based Variable Selection

Suppose the observed data \( x_{ai} \) which are values of \( n \) objects (samples) with respect to \( p \) variables (dimensions) are denoted by the following:

\[
X = (x_{ai}), \quad i = 1, \ldots, n, \quad a = 1, \ldots, p.
\]

(1)

We discuss data when \( p \) is much larger than \( n \), often written \( p >> n \). This data is supposed to have an external information for classification that as data is labeled into \( \hat{K} \) classes. The labeled data are shown as follows:

\[
X_{\hat{k}} = (x_{ai}), \quad i = 1, \ldots, n_{\hat{k}}, \quad a = 1, \ldots, p, \quad \hat{k} = 1, \ldots, \hat{K},
\]

(2)

where \( \sum_{\hat{k}=1}^{\hat{K}} n_{\hat{k}} = n \). Objects in \( X \) is ordered according to the label’s order. We propose a variable selection criterion to reduce the number of variables based on the external information of the classification as follows:

\[
C(a) = \frac{1}{n} \sum_{i=1}^{n} \sum_{a_{k}a} u_{kia} + \cdots + \sum_{a_{k}a} u_{\hat{k}ka}, \quad a = 1, \ldots, p,
\]

(3)

where \( u_{kia} \) shows degree of belongingness of an object \( i_{\hat{k}} \) to a class \( \hat{k} \) with respect to a variable \( a \). The object \( i_{\hat{k}} \) corresponds to an object labeled to a class \( \hat{k} \) which is represented as \( x_{ij} = (x_{i1j}, \ldots, x_{ipj})^{T} \) in equation (2). \( u_{kia} \) is assumed to satisfy the following conditions:

\[
u_{kia} \in [0, 1], \quad \forall i_{\hat{k}}, \hat{k}, a, \quad \sum_{\hat{k}=1}^{\hat{K}} u_{kia} = 1, \quad \forall i_{\hat{k}}, a.
\]

(4)
From equation (4), the criterion shown in equation (3) can show how the obtained classification structure at each variable adjusts to the given external classification structure and \(0 \leq C(a) \leq 1\). The larger value of \(C(a)\) shows the greater explanatory power for the external classification information. Therefore, using a threshold for \(C(a)\), we can select variables capable of explaining the external classification information of data. In order to obtain the clustering results \(u_{ik},\) we use a fuzzy clustering. We use the \(u_{ik}\) as a general notation. Suppose \(d_{ij}\) is \((i, j)\)-th element of a distance matrix \(D_a\) and shows dissimilarity between objects \(i\) and \(j\) with respect to a variable \(a\). This is defined as follows:

\[
d_a = (d_{ij}), \quad d_{ij} = \sqrt{(x_{ai} - x_{aj})^2}, \quad i, j = 1, \cdots, n, \quad a = 1, \cdots, p.
\]  

(5)

For the fuzzy clustering method in which the target data is dissimilarity data, the fanny method [5] is used. The objective function of this method is defined as follows:

\[
J(\bar{U}) = \sum_{k=1}^{K} \left[ \sum_{i=1}^{n} \frac{1}{s_k} \left( \sum_{j=1}^{n} (\bar{u}_{ik} - \bar{y}_{ik})^m \right)^2 / 2 \sum_{i=1}^{n} (\bar{u}_{ik})^m \right].
\]

(6)

Where, \(\bar{u}_{ik}\) shows degree of belongingness of an object \(i\) to a cluster \(k\) and satisfies the conditions \(\bar{u}_{ik} \in [0, 1], \forall i, k, \sum_{k=1}^{K} \bar{u}_{ik} = 1, \forall i, m, (1 < m < \infty)\) shows a control parameter which can control fuzziness of the belongingness. \(d_{ij}\) shows dissimilarity between objects \(i\) and \(j\). The purpose of this method is to estimate \(\bar{U} = (\bar{u}_{ik})\) which minimize equation (6). In equation (6), the objective function with respect to a variable \(a\) is redefined by using (5) as follows [7]:

\[
J(U_a) = \sum_{k=1}^{K} \left[ \sum_{i=1}^{n} \sum_{j=1}^{n} (u_{ik})^m (u_{jk})^m d_{ij} / 2 \sum_{i=1}^{n} (u_{ik})^m \right], \quad a = 1, \cdots, p.
\]

(7)

Where \(U_a\), \((a = 1, \cdots, p)\) is a matrix for \(a\)-th variable whose element \(u_{ik}\) shows degree of belongingness of an object \(i\) to a cluster \(k\) with respect to a variable \(a\). \(u_{ik}\) can be estimated by minimizing equation (7) under the conditions \(u_{ik} \in [0, 1], \forall i, k, a, \sum_{k=1}^{K} u_{ik} = 1, \forall i, a\).

3. Variable Reduction by Clustering Interval Variables to Categories based on External Classification

If the number of variables \(p\) is extremely large when compared with the number of objects \(n\) \((p >> n)\), then the variable selection has a problem; when the threshold value for \(C(a)\) is large, loss of the data information will be large, consequently the selected variables are not sufficient to explain the data structure. Likewise, when the threshold value for \(C(a)\) is small, then still we have a problem of \(p > n\). In order to solve this problem, we propose a method to summarize the selected variables and transform the remained data after the variable selection to form data as \(p < n\). First, the summarization of variables is stated in this section. Suppose the remained data after the variable selection shown in the previous section as follows:

\[
\tilde{X} = (\tilde{x}_{ik}), \quad i = 1, \cdots, n, \quad \tilde{a} = 1, \cdots, \bar{p},
\]

(8)

where \(\bar{p} < p\), however it is still \(\bar{p} > n\). First, we transform the data to include the external classification information of objects. We use interval to represent each class with respect to each variable as follows:

\[
Y = (y_{\tilde{a}k}) = ([y_{\tilde{a}k}, \bar{y}_{\tilde{a}k}]), \quad \tilde{a} = 1, \cdots, \bar{p}, \quad \tilde{k} = 1, \cdots, \bar{K},
\]

(9)

where \(y_{\tilde{a}k} = [\bar{y}_{\tilde{a}k}, \underline{y}_{\tilde{a}k}]\) shows the interval-valued data of the \(\tilde{a}\)-th variable with respect to a class \(\tilde{k}\) which has the minimum value \(\underline{y}_{\tilde{a}k}\) and the maximum value \(\bar{y}_{\tilde{a}k}\). From equations (2) and (8), \(y_{\tilde{a}k}\) and \(\bar{y}_{\tilde{a}k}\) are obtained as follows:

\[
y_{\tilde{a}k} = \min_{i} \tilde{x}_{i\tilde{a}k}, \quad \bar{y}_{\tilde{a}k} = \max_{i} \tilde{x}_{i\tilde{a}k}, \quad \tilde{a} = 1, \cdots, \bar{p}.
\]

(10)

Equation (10) means that \(\bar{K}\) classes over the objects which is given as external classification information are expressed by \(\bar{K}\) intervals. Since the purpose of this study is identifying a subspace spanned by variables so that the subspace has strong discriminative power adjusted for the externally given classification structure of data, we assume that
the given classification structure has a well separated structure. That is, we do not consider the outliers of data for classes. Although the interval representation of data is sometimes sensitive for the outliers of data, this is the reason why we can use the interval representation to the classes based on the given classification structure. In order to obtain the similarity structure of variables over the \( \hat{K} \) classified objects, we classify the data shown in equation (9). The dissimilarity between \( y_h = (y_{h1}, \ldots, y_{hk}) \) and \( y_b = (y_{b1}, \ldots, y_{bk}) \) is defined as \( d_{ab} = \sum_{k=1}^{\hat{K}} \text{sup}(d(x, y_{h_k})|x \in y_{a_k}) \), \( d(x, y_{h_k}) = \text{inf}(d(x, y)|y \in y_{h_k}) \) and \( d_{ba} = \sum_{k=1}^{\hat{K}} \text{sup}(d(y_{b_k}, y)|y \in y_{a_k}) \), \( d(y_{b_k}, y) = \text{inf}(d(x, y)|x \in y_{a_k}) \). Where, \( d(x, y) \) shows distance between \( x \) and \( y \), \( \forall x \in y_{a_k}, \forall y \in y_{b_k} \). Therefore, \( d_{ab} \neq d_{ba}, (a \neq b) \). We use the symmetric part of the dissimilarity as follows: \( \tilde{d}_{ab} = (d_{ab} + d_{ba})/2 \). Applying this dissimilarity \( \tilde{d}_{ab} \) to the fanny method shown in equation (6), we obtain a fuzzy clustering result

\[
\tilde{U} = (\tilde{a}_{ik}), \; \tilde{a} = 1, \ldots, \tilde{p}, \; \tilde{k} = 1, \ldots, \hat{K}, \quad (11)
\]

under the conditions \( \tilde{u}_{ik} \in [0, 1], \; \forall \tilde{a}, \tilde{k} \), \( \sum_{i=1}^{\tilde{p}} \tilde{u}_{ik} = 1, \; \forall \tilde{a}, \hat{K} \) is a number of categories satisfied \( \hat{K} < n \). Based on the result shown in equation (11), the data shown in equation (8) is categorized into \( \hat{K} \) categories as follows:

\[
\tilde{X}_k = \{ x_{a1} | \tilde{p}_{a1} = 1 \}, \; \tilde{x}_{a1} = (\tilde{x}_{a11}, \ldots, \tilde{x}_{a1n}), \; \tilde{k} = 1, \ldots, \hat{K}, \quad (12)
\]

where \( \tilde{p}_{a1} \) satisfy \( \tilde{u}_{ik} = \text{max}_{1 \leq k \leq \hat{K}} \tilde{u}_{ik} \rightarrow \tilde{p}_{a1} = 1, \; \tilde{a} = 1, \ldots, \tilde{p}, \hat{k} \) under the condition of \( \sum_{k=1}^{\hat{K}} \tilde{p}_{a1} = 1 \). In the case that \( \max_{1 \leq k \leq \hat{K}} \tilde{u}_{ik} \) is not unique, we select the first category which appears that has the maximum degree of belongingness over the categories. We rewrite the data sets \( \tilde{X}_k \) shown in equation (12) as follows:

\[
\tilde{X}_k = (\tilde{x}_{a1j}), \; i = 1, \ldots, n, \; \tilde{a}_k = 1, \ldots, \tilde{p}_k, \; \tilde{k} = 1, \ldots, \hat{K}, \quad (13)
\]

where \( \sum_{k=1}^{\hat{K}} \tilde{p}_k = \tilde{p} \).

4. Clustering Interval Objects to Clusters based on Classification of Variables

In order to create the \( \hat{K} < n \) type data, variables included to the same category is summarized for a fixed object by using an interval as follows:

\[
\tilde{Y} = (\tilde{y}_{ik}) = ((\tilde{y}_{ik1}, \tilde{y}_{ik}), \; i = 1, \ldots, n, \; \tilde{k} = 1, \ldots, \hat{K}, \quad (14)
\]

where \( \tilde{y}_{ik} = (\tilde{y}_{ik1}, \tilde{y}_{ik}) \) shows the interval-valued data of the \( i \)-th object with respect to a category \( \tilde{k} \) which has the minimum value \( \tilde{y}_{ik1} \) and the maximum value \( \tilde{y}_{ik} \). From equation (13), \( \tilde{y}_{ik1} \) and \( \tilde{y}_{ik} \) are obtained as follows:

\[
\begin{align*}
\tilde{y}_{ik1} &= \min_{a_{ik}} \tilde{x}_{ik1}, \\
\tilde{y}_{ik} &= \max_{a_{ik}} \tilde{x}_{ik1}, \\
&i = 1, \ldots, n.
\end{align*}
\]

Equation (15) shows that uncertainty of variables for a category with respect to a fixed object is represented by an interval. Since the categories are obtained by using a clustering method, we can assume that the obtained classification structure of variables has a well separated structure. Therefore, we use the interval representation for each category with respect to an object without consideration of outliers of data. Since \( \hat{K} < n \) in equation (14), we can apply this data to a clustering method shown in equation (6) in order to classify the objects. First, we calculate the dissimilarity between objects \( \tilde{y}_i = (\tilde{y}_{i1}, \ldots, \tilde{y}_{ik}) \) and \( \tilde{y}_j = (\tilde{y}_{j1}, \ldots, \tilde{y}_{jk}) \) as follows:

\[
\begin{align*}
d_{ij} &= \sum_{k=1}^{\hat{K}} \text{sup}(d(x, \tilde{y}_{ik})|x \in \tilde{y}_{ik}), \\
&d(x, \tilde{y}_{ik}) = \text{inf}(d(x, y)|y \in \tilde{y}_{ik}), \\
d_{ji} &= \sum_{k=1}^{\hat{K}} \text{sup}(d(\tilde{y}_{ik}, y)|y \in \tilde{y}_{ik}), \\
&d(\tilde{y}_{ik}, y) = \text{inf}(d(x, y)|x \in \tilde{y}_{ik}).
\end{align*}
\]

Where, \( d(x, y) \) shows distance between \( x \) and \( y \), \( \forall x \in \tilde{y}_{ik}, \forall y \in \tilde{y}_{ik} \). Therefore, \( d_{ij} \neq d_{ji}, (i \neq j) \). We use the symmetric part of the dissimilarity as follows: \( \tilde{d}_{ij} = (d_{ij} + d_{ji})/2 \). Applying this dissimilarity \( \tilde{d}_{ij} \) to the fanny method shown in equation (6), we obtain a fuzzy clustering result

\[
\tilde{U} = (\tilde{u}_{ik}), \; i = 1, \ldots, n, \; k = 1, \ldots, \hat{K}, \quad (16)
\]
under the conditions,
\[ \tilde{u}_{ik} \in [0, 1], \forall i, k, \sum_{k=1}^{K} \tilde{u}_{ik} = 1, \forall i, \] (19)
where \( K \) is a number of clusters satisfied \( K < n \).

5. Selection of number of clusters

Since according to the change of number of clusters, the obtained classification structure is changed, obtaining an adaptable classification structure which can represent the dissimilarity structure well is closely related with how to determine an adaptable number of clusters. The criterion of selection of an appropriate number of clusters is defined as follows [8]:
\[ C(K) = \sum_{i \neq j=1}^{n} s_{ij}^{(K)} / \left( \sqrt{\sum_{i \neq j=1}^{n} s_{ij}^{(K)} \sum_{i \neq j=1}^{n} s_{ij}^{(K)}} \right), \] (20)
where \( s_{ij} \) shows a similarity between objects \( i \) and \( j \) and is calculated from data shown in equation (14) as follows:
\[ S = (s_{ij}), \quad s_{ij} = 1 - d_{ij}/\max_{i,j}(d_{ij}), \quad i, j = 1, \ldots, n. \] The dissimilarity \( d_{ij} \) between \( \tilde{y}_i = (\tilde{y}_{i1}, \ldots, \tilde{y}_{ik}) \) and \( \tilde{y}_j = (\tilde{y}_{j1}, \ldots, \tilde{y}_{jk}) \) is obtained by using equations (16) and (17). \( s_{ij}^{(K)} \) shows the restored similarity obtained as follows [6]:
\[ s_{ij}^{(K)} = \sum_{k=1}^{K} \sum_{l=1}^{K} w_{kl}^{(K)} u_{ik}^{(K)} u_{jl}^{(K)}, \] (21)
where \( w_{kl}^{(K)} \) is considered to be a quantity which shows the asymmetric similarity between a pair of clusters when we assume the number of clusters as \( K \). In this paper, we define the \( w_{kl}^{(K)} \) as derived from an assumption of normal distribution of objects in each cluster as follows:
\[ w_{kl}^{(K)} = 1 - 1/(1 + e^{-\alpha_{kl}^{(K)}}, \quad \alpha_{kl}^{(K)} = 1/2 \left( \|\mu_{ik} - \mu_{lk}\|_{\Sigma_{ik}^{-1}} + tr(\Sigma_{ik}^{-1} \Sigma_{lk} - I) \right. \) + \log \left( \sqrt{\Sigma_{ik}^{-1} \Sigma_{lk}} \right), \] (22)
where \( \|\mu_{ik} - \mu_{lk}\|_{\Sigma_{ik}^{-1}} \) shows the similarity from a cluster \( k \) to a cluster \( l \) when we assume the number of clusters is \( K \). \( I \) is a unit matrix. \( \mu_{ik}, \Sigma_{ik}, \mu_{lk}, \Sigma_{lk} \) are an expected value and a variance-covariance matrix of \( S_{ikl} \) which is shown as follows:
\[ S_{ikl} = (\tilde{y}_i | p_{ik}^{(K)} = 1), \quad \tilde{y}_i = (\tilde{y}_{i1}, \ldots, \tilde{y}_{ip}), \quad \tilde{y}_{ia} = (\tilde{y}_{ia} + \tilde{y}_{ia})/2, \] \( \forall k \), where \( p_{ik}^{(K)} \) satisfy
\[ u_{ik}^{(K)} = \max_{1 \leq l \leq K} u_{ik}^{(K)} \rightarrow p_{ik}^{(K)} = 1, \quad i = 1, \ldots, n, \] under the condition of \( \sum_{k=1}^{K} p_{ik}^{(K)} = 1 \). \( u_{ik}^{(K)} \) shows degree of belongingness of an object \( i \) to a cluster \( k \) when we assume the number of clusters is \( K \), and satisfy the condition (19). \( u_{ik}^{(K)} \) is obtained by applying calculated symmetrized dissimilarity derived from equations (16) and (17) to a fuzzy clustering shown in equation (6). From equation (22), \( w_{kl}^{(K)} \neq w_{lk}^{(K)}, \quad (k \neq l) \), \( w_{kl}^{(K)} \in [0, 1] \) are satisfied. \( C(K) \) shown in equation (20) shows the degree of alignment between \( s_{ij} \) and \( s_{ij}^{(K)} \). Therefore, the larger value of \( C(K) \) is better when compared with several cases in which we assume several numbers of clusters shown as \( K \). In other words, selecting the best \( K \) when we obtain the largest value of \( C(K) \) means selecting the best matched latent classification structure of original similarity matrix, \( S = (s_{ij}) \), since \( s_{ij}^{(K)} \) shown in equation (21) involves the latent classification structure of \( s_{ij} \) when the number of clusters is fixed as \( K \). The concentration around the expected value of the criterion shown in equation (20) for the different \( w_{kl}^{(K)} \) has been proven. [8]

6. Numerical example

We use gene expression data for prostate cancer. [9] The data consists of 32 objects (subjects) with respect to 12626 variables (genes) shown in equation (1). As external classification information, 32 objects are labeled into two clusters of which 23 objects are from shavings of prostate tissue with cancer and 9 objects from shavings of prostate tissue are without cancer. The purpose is to identify variables (genes) and obtain the categories of variables (genes)
which explain the classification structure of the two classes (a class of 23 objects with cancer and a class of 9 objects without cancer).

Using the variable selection criterion shown in equation (3), we obtained values of the criterion for each variable (gene). Figure 1 shows frequency distribution of the criterion values. Figure 2 shows variance of values of the criterion for each range. From figures 1 and 2, we can see that variance of the values which are larger than 0.8 is significantly smaller when compared with other ranges. This means that robustness for the selection of the threshold value is strong when we select the values which are larger than 0.8. Therefore, we selected variables which have more than 0.8 for the criterion. 90 variables (genes) are selected. Based on the classification of variables described in section 3, we obtained 6 categories from 90 variables. Using the transformed $32 \times 6$ interval-valued data shown in equation (14), we check the criterion shown in equation (20) when the number of clusters are 2, 3, and 4. The value of $C(2)$ is largest when compared with other two values of $C(3)$ and $C(4)$, so we select the number of clusters as 2. In order to check the classification ability, figure 3 shows the result of proposed clustering method shown in equation (18). In this figure, objects 1-23 show 23 objects from shavings of prostate tissue with cancer and objects 24-32 show 9 objects from shavings of prostate tissue without cancer. The value of ordinate shows the degree of belongingness of objects to each cluster. From this figure, it can be seen that the proposed clustering method successfully classified the two classes. This result can be used for the prediction of a new object. Since we know the selected 90 variables are effective for the discrimination of two classes, we just need to observe the values of the 90 variables (genes). According to the obtained 6 categories of the 90 variables, we create the interval-valued data for the new object using equation (14). Adding the newly obtained interval-valued data to the original data set and applying it to the clustering method shown in section 4, we can obtain the result of fuzzy clustering shown in equation (18) for the new object. From this result, we can discriminate to which classes this object belongs.

7. Conclusion

For high dimension low sample-size data, we propose a variable selection criterion based on individual classification structure on each variable and an identification method for categories of variables for creating a classifier. In order to solve problems of clustering high dimension low sample-size data, which is caused by noise and a lack of robustness of the solution, we first exploit the representation of interval-valued data to overcome the noise problem and then exploit fuzzy clustering to obtain a robust result. A numerical example shows an efficient performance of the proposed clustering method to create the classifier.