Computing and Informatics, Vol. 33, 2014, 131–153

ROUGH FUZZY SUBSPACE CLUSTERING FOR DATA WITH MISSING VALUES

Krzysztof Simiński

Institute of Informatics, Silesian University of Technology
ul. Akademicka 16, 44-100 Gliwice, Poland
e-mail: Krzysztof.Siminski@polsl.pl

Abstract. The paper presents rough fuzzy subspace clustering algorithm and experimental results of clustering. In this algorithm three approaches for handling missing values are used: marginalisation, imputation and rough sets. The algorithm also assigns weights to attributes in each cluster; this leads to subspace clustering. The parameters of clusters are elaborated in the iterative procedure based on minimising of criterion function. The crucial parameter of the proposed algorithm is the parameter having the influence on the sharpness of elaborated subspace cluster. The lower values of the parameter lead to selection of the most important attribute. The higher values create clusters in the global space, not in subspaces. The paper is accompanied by results of clustering of synthetic and real life data sets.

Keywords: Rough fuzzy subspace clustering, clustering, rough set, marginalisation, imputation, missing values

Mathematics Subject Classification 2010: 91C20 (Clustering), 62H30 (Classification and discrimination; cluster analysis), 68T37 (Reasoning under uncertainty)

1 INTRODUCTION

The attributes in data sets have not always the same importance. Sometimes some attributes are of minor importance or even represent nothing but noise. The methods for elimination of useless dimension have been elaborated; but these methods remove attributes globally from all data tuples. The global approach by feature transformation (e.g. PCA or SVD) leads to problems with interpretability of elaborated models. The global elimination of dimensions may not be satisfactory because clusters may exist in different subspaces. This leads to subspace clustering where each cluster may exist in its own subspace [14, 17, 36].

There are two kinds of subspace clustering: bottom-up and top-down [36]. The first approach splits the clustering space with a grid and analyses the density of data examples in each grid cell extracting the relevant attributes/dimensions (e.g. CLIQUE [4], ENCLUS [7], MAFIA [19]). The latter (top-down) approach starts with full dimensional clusters and tries to throw away the dimensions of minor importance (e.g. PROCLUS [2], ORCLUS [3], δ-Clusters [51], FSC [17, 16]).

The next quite often problem is the lack of values in the data tuples. The reasons are various: errors in answer acquisition, impossibility to get data (e.g. patient has died), refusal to answer some questions in the questionnaire, inapplicability of questions, irrelevant or unknown attributes, random noise, impossible values, retrospective usage of data – i.e. the data were gathered for other purpose than the research needs. In [38] a medical example is given where only 1 patient in 55 had all blood tests done. Overall 9.2 % of blood test results are missing. In [28] the real life data set is presented with more than 50 % of values missing. The paper [30] presents three classes of missing data types:

MCAR – missing completely at random: The probability of a tuple having a missing value for an attribute depends neither on the known values nor on the missing data.

MAR – missing at random: The probability that the tuple has a missing value for an attribute may depend on the known values, but not on the value of the missing data itself.

NMAR – not missing at random: The probability of an instance having missing value for an attribute could depend on the value of that attribute.

The paper [1] classifies the difficulty of analysis of data sets with missing values. The data sets with less than 1 % of missing values are labelled as trivial, 1–5 % as manageable. For data sets with 5–15 % of missing values some sophisticated methods are required and finally more than 15 % missing values "severely impact any kind of interpretation".

Generally three approaches are used to handle the problem of missing values:

1. imputation – the unknown values are substituted with estimated ones [38, 48, 10, 18, 53, 54];
2. marginalisation (WDS: whole data strategy) in done in two ways:
   (a) the data tuples with missing values are removed from the data set [46, 23] or
   (b) the features (attributes) with missing values are ignored [8] – this leads to lowering the dimensionality of the task;

3. applying rough sets for expressing the imprecision caused by lack of data [35, 22, 20].

The advantage of both data imputation and marginalisation is in their simplicity. Imputation is more frequently used than marginalisation [24]. The results elaborated based on data sets with imputed values cannot be fully trusted [46]. The imputed values may have no physical meaning in real life [48]. The missing values are commonly replaced with zeros, random numbers, mean value over all data set [33], mean value over the class the example belongs to [21], median imputation, deductive imputation (the missing values are deduced from other information of the pattern), regression-based imputation [6] or value based on real distribution (the missing values are replaced with random values with data set distribution) [48]. Expectation-Maximisation (EM) [10] algorithm is applied in [18]. Imputation based on nearest neighbourhood is proposed in [53, 54], K-Nearest neighbour imputation in [5]. Imputation with fuzzy K-means clustering (FKMI) is described in [1]. To avoid imputation of non existing values the hot-deck procedure has been proposed [39] with various distance measures [15, 11]. Similar to hot-deck approach called cold-deck approach was proposed in [27]. In [31] the Event Covering approach [49] is claimed to be the most suitable for radial basis function network classifiers. The impact of imputation of missing values on classification error is discussed in [12]. The review of imputation methods can be found in [31, 32].

The method proposed by [47, 48] divides the feature (attribute) set $\mathbb{A}$ into features $\mathbb{A}_o$ with no lacking values and partially observed features $\mathbb{A}_m$. Then a special clustering algorithm is applied.

The novelty of the paper is the subspace clustering algorithm for data set with missing values. The algorithm can handle missing value data sets simultaneously by applying imputation, marginalisation and rough sets. This leads to rough fuzzy clustering [44, 43]. The algorithm also assigns the weights to the attributes separately in each cluster [42, 45]. Combining these features leads to rough fuzzy subspace clustering. The clustering is one of the initial steps in fuzzy model creation in neuro-fuzzy system [41, 34, 26, 50]. One of the applications of this algorithm may be the partition of input domain for neuro-fuzzy systems.

The paper is organised as follows: Section 2.1 describes the preprocessing of data (imputation and marginalisation). Section 2.2 introduces a new algorithm for rough fuzzy subspace clustering. Section 3 describes the experiments and finally Section 4 summarizes the conclusions.

In the paper the empty characters ($\mathbb{A}$) are used to denote the sets, bolds ($\mathbf{a}$) – matrices and vectors, uppercase italics (A) – the cardinality of sets, lowercase italics (a) – scalars and set elements. Overline $\overline{a}$ denotes upper value of the rough set, underline $\underline{a}$ – the lower one. Tilde over symbol $\widetilde{a}$ means that the value was calculated with the "upper" data set $\widetilde{\mathbb{X}}$, tilde below a symbol $\underset{\sim}{a}$ stands for the value elaborated with "lower" data set $\underset{\sim}{\mathbb{X}}$ (the sets $\widetilde{\mathbb{X}}$ and $\underset{\sim}{\mathbb{X}}$ are described in Section 2). Detailed list of symbols is given in Table 1.

$\mathbb{R}$	set of clusters
r	cluster, $r \in \mathbb{R}$
R	number of clusters, $R = \Vert\mathbb{R}\Vert$
$\mathbb{X}$	set of tuples, data examples
$\widetilde{\mathbb{X}}$	imputed set of data examples
$\underset{\sim}{\mathbb{X}}$	marginalised set of data examples
$\mathbf{x}$	tuple, data example, $\mathbf{x} \in \mathbb{X}$
$\widetilde{\mathbf{x}}$	tuple from imputed data set, $\mathbf{x} \in \mathbb{X}$
$\underset{\sim}{\mathbf{x}}$	tuple from marginalised data set, $\mathbf{x} \in \mathbb{X}$
$\mathbf{x}_i$	i^{th} tuple
x	descriptor of a tuple, $\mathbf{x} = [x_1, \ldots, x_A]^{\mathrm{T}}$
X	number of tuples, $X = \Vert\mathbb{X}\Vert$
X_u	number of tuples in upper set, $X_u = \left\Vert\widetilde{\mathbb{X}}\right\Vert$
X_l	number of tuples in lower set, $X_l = \left\Vert\underset{\sim}{\mathbb{X}}\right\Vert$
$\mathbb{A}$	set of attributes
a	attribute, $a \in \mathbb{A}$
A	number of attributes in a tuple, $A = \Vert\mathbb{A}\Vert$
A_t	threshold number of attributes
$\boldsymbol{\mu}$	partition matrix
$\widetilde{\mu}_{ru}$	membership value of the u^{th} tuple to r^{th} "upper" cluster
$\underset{\sim}{\mu}_{rl}$	membership value of the l^{th} tuple to r^{th} "lower" cluster
d_{rj}	distance between r^{th} cluster's centre and j^{th} tuple
η_u	weight of the u^{th} tuple in $\widetilde{\mathbb{X}}$
$\mathbf{v}_r$	the centre of r^{th} cluster
$\widetilde{\mathbf{s}}$	the fuzzyfication of the "upper" cluster
$\underset{\sim}{\mathbf{s}}$	the fuzzyfication of the "lower" cluster
$\bar{\mathbf{s}}$	the upper fuzzyfication of the cluster
$\underline{\mathbf{s}}$	the lower fuzzyfication of the cluster
$\mathbf{z}$	vector of weights of the attributes
f	the fuzzyfication parameter

Table 1. Symbols used in the paper

2 OUR APPROACH

The drawback of imputing for handling missing values mentioned in the Introduction is no distinction between original untouched data and imputed values [24]. Further, the imputed values may have no medical/physical meaning [48], thus the models based on imputed data cannot be fully trusted [46].

Our solution applies marginalisation imputation and rough sets. Marginalisation removes the tuples with missing values. The remaining tuples contain only original data. This data set is denoted as $\underset{\sim}{\mathbb{X}}$. Imputation is used to handle data with missing values. All data augmented with imputed data stand for upper data set $\widetilde{\mathbb{X}}$. The lower data set is a subset of upper data set: $\underset{\sim}{\mathbb{X}} \subseteq \widetilde{\mathbb{X}}$. Both sets $\underset{\sim}{\mathbb{X}}$ and $\widetilde{\mathbb{X}}$ are used to elaborate the rough set clusters. This approach maintains the distinction between original and imputed values. If the data set lacks no values the "upper" and "lower" data sets are the same: $\underset{\sim}{\mathbb{X}} = \widetilde{\mathbb{X}} = \mathbb{X}$.

2.1 Preprocessing of Data

Preprocessing of data creates $\underset{\sim}{\mathbb{X}}$ and $\widetilde{\mathbb{X}}$ data sets from original data set $\mathbb{X}$. The first one is created with marginalisation, the latter with imputation. The following sections describe these procedures in detail.

2.1.1 Marginalisation

The data set $\underset{\sim}{\mathbb{X}}$ contains only these data tuples from $\mathbb{X}$ that lack no values. Marginalisation excludes the tuples, not the attributes, thus there is no dimensionality reduction. This approach is similar to that used by [46, 23].

2.1.2 Imputation

The tuples with missing values are substituted with new tuples with imputed values. If the tuple lacks n values, it is substituted with k^n tuples with all combination of imputed values (these are the mean values m of missing attribute calculated from values existing in other tuples, $m + \sigma$, where σ is the standard deviation of the attribute, $m - \sigma$, thus $k = 3$). The maximum and minimum values are not used here, because the extreme values may be outliers and one extreme bias value can substantially influence the clustering process. This approach will be later referenced as full imputation. Unfortunately the number of new tuples grows very fast with the number n of missing values from the original tuple. This explosion in number of tuples can have disadvantageous influence on the efficacy of calculations. Thus when the tuple lacks more than A_t values not all possible combinations are imputed, but for each missing value v only k new tuples are created and other missing attributes $q \neq v$ are imputed by means of the respective attributes. So only kn new tuples are added. Figure 1 presents an example of a data set with missing values denoted with question marks. If $A_t \geqslant 2$ the tuple $\mathbf{x}_1$ will be substituted with $k^n = 3^2 = 9$ tuples – Figure 2. If $A < 2$ the tuple in question will be imputed with $kn = 3 \cdot 2 = 6$ tuples – Figure 3. The twofold approach is used because in real-life data set the tuple may lack 8 or more values.

If the tuple with missing values is substituted with t imputed tuples, each of these imputed tuples is assigned weight $\eta = 1/t$. The weight is treated as condition in conditional clustering (Equation (3)).

	a_1	a_2	a_3	a_4
$\mathbf{x}_1$	2	?	?	1
$\mathbf{x}_2$	1	4	6	3
$\mathbf{x}_3$	8	5	7	3
$\mathbf{x}_4$	5	2	9	1
$\mathbf{x}_5$	8	3	6	2
$\mathbf{x}_6$	7	0	5	4
average		2.80	6.60	
st. dev.		1.72	1.36	

Figure 1. Example of the data set with missing values (denoted with question marks). The last two rows show the average values and standard deviation of attributes a_2 and a_3.

	a_1	a_2	a_3	a_4
$\mathbf{x}_1$	2	?	?	1
$\mathbf{x}_7$	2	1.08	5.24	1
$\mathbf{x}_8$	2	1.08	6.60	1
$\mathbf{x}_9$	2	1.08	7.96	1
$\mathbf{x}_{10}$	2	2.80	5.24	1
$\mathbf{x}_{11}$	2	2.80	6.60	1
$\mathbf{x}_{12}$	2	2.80	7.96	1
$\mathbf{x}_{13}$	2	4.52	5.24	1
$\mathbf{x}_{14}$	2	4.52	6.60	1
$\mathbf{x}_{15}$	2	4.52	7.96	1

Figure 2. Tuples No. 7–15 are imputed in the data set from Figure 1 in place of tuple $\mathbf{x}_1$ when $A_t \geqslant n = 2$

2.2 Rough-Fuzzy Subspace Clustering

The clustering method is based on minimising the criterion function:

$$J = \sum_{r=1}^{R} \left[\sum_{u=1}^{X_u} \widetilde{\mu}_{ru}^{m} \sum_{a=1}^{A} z_{ra}^{f} d_{ra}^{2} \left(\widetilde{\mathbf{x}}_u\right) + \sum_{l=1}^{X_l} \underset{\sim}{\mu}_{rl}^{m} \sum_{a=1}^{A} z_{ra}^{f} d_{ra}^{2} \left(\underset{\sim}{\mathbf{x}}_l\right) \right], \tag{1}$$

where

$$d_{ra}^{2} \left(\mathbf{x}\right) = \left(x_{ka} - v_{ra}\right)^2 \tag{2}$$

is the distance for the a^{th} attribute from the r^{th} cluster centre for the tuple $\mathbf{x}$. The symbol $\widetilde{\mathbf{x}}$ denotes the tuple form the "upper" data set $\widetilde{\mathbb{X}}$ and $\underset{\sim}{\mathbf{x}}$ is a tuple from "lower" data set $\underset{\sim}{\mathbb{X}}$.

	a_1	a_2	a_3	a_4
$\mathbf{x}_1$	2	?	?	1
$\mathbf{x}_{16}$	2	2.80	5.24	1
$\mathbf{x}_{17}$	2	2.80	6.60	1
$\mathbf{x}_{18}$	2	2.80	7.96	1
$\mathbf{x}_{19}$	2	1.08	6.60	1
$\mathbf{x}_{20}$	2	2.80	6.60	1
$\mathbf{x}_{21}$	2	4.52	6.60	1

Figure 3. Tuples No. 16–21 are imputed in the data set from Figure 1 in place of tuple $\mathbf{x}_1$ when $A_t < n = 2$

For each data tuple from "upper" data set the conditional boundary is

$$\underset{u \in \widetilde{\mathbb{X}}}{\forall} \sum_{r=1}^{R} \widetilde{\mu}_{ru} = \eta_u, \tag{3}$$

where η_u is a tuple's weight (cf. Section 2.1.2). This condition is similar to the condition used in conditional FCM algorithm [37]. For lower clustering the standard FCM boundary is applied:

$$\underset{l \in \underset{\sim}{\mathbb{X}}}{\forall} \sum_{r=1}^{R} \underset{\sim}{\mu}_{rl} = 1. \tag{4}$$

The cluster centres are elaborated based only on "lower" membership values:

$$\mathbf{v}_i = \frac{\sum_{l=1}^{X_l} \underset{\sim}{\mu}_{il} \, \underset{\sim}{\mathbf{x}}_l}{\sum_{l=1}^{X_l} \underset{\sim}{\mu}_{il}}. \tag{5}$$

Constraints for dimension weights $\mathbf{z}$ are similar:

$$\underset{r \in \mathbb{R}}{\forall} \sum_{a=1}^{A} z_{ra} = 1. \tag{6}$$

The Lagrange function for criterion function can be expressed as

$$\begin{aligned} L(\boldsymbol{\mu}, \lambda_1, \lambda_2, \lambda_3) \;=\; & \sum_{r=1}^{R} \left[\sum_{u=1}^{X_u} \widetilde{\mu}_{ru}^{m} \sum_{a=1}^{A} z_{ra}^{f} d_{ra}^{2} \left(\widetilde{\mathbf{x}}_u \right) + \sum_{l=1}^{X_l} \underset{\sim}{\mu}_{rl}^{m} \sum_{a=1}^{A} z_{ra}^{f} d_{ra}^{2} \left(\underset{\sim}{\mathbf{x}}_l \right) \right] \\ & - \lambda_1 \sum_{r=1}^{R} \left[\sum_{a=1}^{A} z_{ra} - 1 \right] \\ & - \lambda_2 \sum_{l=1}^{X_l} \left[\sum_{r=1}^{R} \underset{\sim}{\mu}_{rl} - 1 \right] - \lambda_3 \sum_{u=1}^{X_u} \left[\sum_{r=1}^{R} \widetilde{\mu}_{ru} - w_u \right]. \end{aligned} \tag{7}$$

For the r^{th} rule

$$\frac{\partial L}{\partial \widetilde{\mu}_{ru}} = m\widetilde{\mu}_{ru}^{m-1}\left(\sum_{a=1}^{A} z_{ra}^{f} d_{ra}^{2}\left(\widetilde{\mathbf{x}}_{u}\right)\right) - \lambda_3 = 0 \tag{8}$$

$$\frac{\partial L}{\partial \underset{\sim}{\mu}_{rl}} = m\underset{\sim}{\mu}_{rl}^{m-1}\left(\sum_{a=1}^{A} z_{ra}^{f} d_{ra}^{2}\left(\underset{\sim}{\mathbf{x}}_{l}\right)\right) - \lambda_2 = 0 \tag{9}$$

$$\frac{\partial L}{\partial z_{ra}} = \sum_{l=1}^{X_l} \underset{\sim}{\mu}_{rl}^{m} f\, z_{ra}^{f-1} d_{ra}^{2}\left(\underset{\sim}{\mathbf{x}}_{l}\right) - \lambda_1 = 0. \tag{10}$$

Transforming Equation (8) we get:

$$\frac{\partial L}{\partial \widetilde{\mu}_{ru}} = m\widetilde{\mu}_{ru}^{m-1}\left(\sum_{a=1}^{A} z_{ra}^{f} d_{ra}^{2}\left(\widetilde{\mathbf{x}}_{u}\right)\right) - \lambda_3 = 0, \tag{11}$$

$$\frac{\lambda_3}{m} = \widetilde{\mu}_{ru}^{m-1}\left(\sum_{a=1}^{A} z_{ra}^{f} d_{ra}^{2}\left(\widetilde{\mathbf{x}}_{u}\right)\right), \tag{12}$$

$$\widetilde{\mu}_{ru} = \left(\frac{\lambda_3}{m}\right)^{\frac{1}{m-1}}\left(\sum_{a=1}^{A} z_{ra}^{f} d_{ra}^{2}\left(\widetilde{\mathbf{x}}_{u}\right)\right)^{\frac{1}{1-m}}. \tag{13}$$

Analogously for Equation (9):

$$\underset{\sim}{\mu}_{rl} = \left(\frac{\lambda_2}{m}\right)^{\frac{1}{m-1}}\left(\sum_{a=1}^{A} z_{ra}^{f} d_{ra}^{2}\left(\underset{\sim}{\mathbf{x}}_{l}\right)\right)^{\frac{1}{1-m}}, \tag{14}$$

substituting 13 into 3:

$$w_u = \sum_{r=1}^{R}\left(\frac{\lambda_3}{m}\right)^{\frac{1}{m-1}}\left(\sum_{a=1}^{A} z_{ra}^{f} d_{ra}^{2}\left(\widetilde{\mathbf{x}}_{u}\right)\right)^{\frac{1}{1-m}}. \tag{15}$$

Dividing 13 by 15 gives:

$$\widetilde{\mu}_{ru} = w_u \frac{\left(\sum_{a=1}^{A} z_{ra}^{f} d_{ra}^{2}\left(\widetilde{\mathbf{x}}_{u}\right)\right)^{\frac{1}{1-m}}}{\sum_{r=1}^{R}\left(\sum_{a=1}^{A} z_{ra}^{f} d_{ra}^{2}\left(\widetilde{\mathbf{x}}_{u}\right)\right)^{\frac{1}{1-m}}}. \tag{16}$$

Analogously the $\underset{\sim}{\mu}_{rl}$ is calculated:

$$\underset{\sim}{\mu}_{rl} = \frac{\left(\sum_{a=1}^{A} z_{ra}^{f} d_{ra}^{2}\left(\underset{\sim}{\mathbf{x}}_{l}\right)\right)^{\frac{1}{1-m}}}{\sum_{r=1}^{R}\left(\sum_{a=1}^{A} z_{ra}^{f} d_{ra}^{2}\left(\underset{\sim}{\mathbf{x}}_{l}\right)\right)^{\frac{1}{1-m}}}. \tag{17}$$

Now the dimension weights are to be calculated. When $f \neq 1$ the transformation of Equation (10) gives:

$$\sum_{u=1}^{X_u} \underset{\sim}{\mu}_{rl}^{m} \, f \, z_{ra}^{f-1} d_{ra}^{2} \left(\underset{\sim}{\mathbf{x}}_u\right) - \lambda_1 \;=\; 0 \tag{18}$$

$$\sum_{u=1}^{X_l} \underset{\sim}{\mu}_{rl}^{m} \, f \, z_{ra}^{f-1} d_{ra}^{2} \left(\underset{\sim}{\mathbf{x}}_u\right) \;=\; \lambda_1 \tag{19}$$

$$f \, z_{ra}^{f-1} \sum_{l=1}^{X_l} \underset{\sim}{\mu}_{rl}^{m} \, d_{ra}^{2} \left(\underset{\sim}{\mathbf{x}}_u\right) \;=\; \lambda_1 . \tag{20}$$

Further:

$$z_{ra}^{f-1} \;=\; \frac{\lambda_1}{f \sum_{l=1}^{X_l} \underset{\sim}{\mu}_{rl}^{m} \, d_{ra}^{2} \left(\underset{\sim}{\mathbf{x}}_u\right)} \tag{21}$$

$$z_{ra} \;=\; \left(\frac{\lambda_1}{f}\right)^{\frac{1}{f-1}} \left(\sum_{u=1}^{X_l} \underset{\sim}{\mu}_{ru}^{m} \, d_{ra}^{2} \left(\underset{\sim}{\mathbf{x}}_u\right)\right)^{\frac{1}{1-f}} . \tag{22}$$

Combining Equations (22) and (6) we get:

$$1 = \sum_{a=1}^{A} \left(\frac{\lambda_1}{f}\right)^{\frac{1}{f-1}} \left(\sum_{l=1}^{X_u} \underset{\sim}{\mu}_{rl}^{m} \, d_{ra}^{2} \left(\underset{\sim}{\mathbf{x}}_u\right)\right)^{\frac{1}{1-f}} . \tag{23}$$

Dividing 22 by 23 gives:

$$z_{ra} = \frac{\left(\sum_{l=1}^{X_l} \underset{\sim}{\mu}_{rl}^{m} \, d_{ra}^{2} \left(\underset{\sim}{\mathbf{x}}_u\right)\right)^{\frac{1}{1-f}}}{\sum_{a=1}^{A} \left(\sum_{l=1}^{X_l} \underset{\sim}{\mu}_{rl}^{m} \, d_{ra}^{2} \left(\underset{\sim}{\mathbf{x}}_l\right)\right)^{\frac{1}{1-f}}} . \tag{24}$$

When $f = 1$ somewhat different approach has to be applied. In such a situation the objective function (1) becomes

$$J = \sum_{r=1}^{R} \left[\sum_{u=1}^{X_u} \widetilde{\mu}_{ru}^{m} \sum_{a=1}^{A} z_{ra} d_{ra}^{2} \left(\widetilde{\mathbf{x}}_u\right) + \sum_{l=1}^{X_l} \underset{\sim}{\mu}_{rl}^{m} \sum_{a=1}^{A} z_{ra} d_{ra}^{2} \left(\underset{\sim}{\mathbf{x}}_l\right) \right] . \tag{25}$$

In such a case the attribute a of the r^{th} rule for which the sum

$$\sum_{u=1}^{X_u} \widetilde{\mu}_{ru}^{m} d_{ra}^{2} \left(\widetilde{\mathbf{x}}_u\right) + \sum_{l=1}^{X_l} \underset{\sim}{\mu}_{rl}^{m} d_{ra}^{2} \left(\underset{\sim}{\mathbf{x}}_l\right) \tag{26}$$

is minimal gets the weight $z_{ra} = 1$ and other attributes of this rule get zero weights (because of the constraint expressed by Equation 6).

```
input:  𝕏 {array of tuples}
input:  maxIter {maximal number of iterations}
output: v {clusters' centres}
output: s̄ {fuzzification of ``upper'' rough set}
output: s̲ {fuzzification of ``lower'' rough set}
output: z {attributes' weight matrix}

 {initialization:}
 numberOfIterations := 0;
 {preprocessing of data:}
 𝕏̃ := imputation(𝕏) {creation of ``upper'' data set}
 𝕏̰ := marginalisation(𝕏) {creation of ``lower'' data set}
 {clustering:}
 random initialization of μ̃, μ̰ and z;
 while numberOfIterations < maxIter do
 {
    calculate v; {Equation~5}
    update μ̃, μ̰; {Eqq. 16, 17}
    update z; {Equation~24}
    numberOfIterations := numberOfIterations + 1;
 }
 end while;
 calculate v; {Equation~5}
 augment z; {Equation~27}
 calculate s̃, s̰; {Eqq. 29, 30}
 calculate s̄, s̲; {Eqq. 31, 32}
```

Figure 4. Rough fuzzy subspace clustering

Alternating application of Equations 5, 17, 16 and 24 leads to the algorithm presented in Figure 4.

For clustering procedure the sum weights of attributes in one rule have the constraint expressed by Equation (6). When there are many attributes the values of weights are very small, what makes them difficult for interpretation. For more clarity the weights are augmented when the clustering procedure has been finished. The attribute weights for one rule are divided by their maximal values. This maximal value is always greater than zero. In this procedure all weights in this rule are scaled and the maximum weights become one:

$$\underset{r \in \mathbb{R}}{\forall} z_{ra} \leftarrow \frac{z_{ra}}{\max_{i \in [1..A]} z_{ri}}. \tag{27}$$

2.3 Extraction of Clusters from Partition Matrices

The membership function to clusters is defined as Gaussian function

$$g(x; v, s) = \exp\left(-\frac{(x-v)^2}{2s^2}\right). \tag{28}$$

Based on partition matrices $\widetilde{\boldsymbol{\mu}}$ and $\underset{\sim}{\boldsymbol{\mu}}$ the parameters v, s are calculated. The cluster center $\mathbf{v} = [v_1, v_2, \ldots, v_A]$ is calculated by help of Equation (5). The s parameter is elaborated by help of Equations [9, 29]

$$\widetilde{\mathbf{s}}_i = \sqrt{\frac{\sum_{u=1}^{X_u} \widetilde{\mu}_{iu}^m \left(\widetilde{\mathbf{x}}_u - \mathbf{v}_i\right)}{\sum_{u=1}^{X_u} \widetilde{\mu}_{iu}}} \tag{29}$$

for "upper" clusters and

$$\underset{\sim}{\mathbf{s}}_i = \sqrt{\frac{\sum_{l=1}^{X_l} \underset{\sim}{\mu}_{il}^m \left(\underset{\sim}{\mathbf{x}}_l - \mathbf{v}_i\right)}{\sum_{l=1}^{X_l} \underset{\sim}{\mu}_{il}}} \tag{30}$$

for "lower" ones. It cannot be guaranteed that the values $\underset{\sim}{\mathbf{s}}$ elaborated with "lower" data set are $\underline{\mathbf{s}}$ and $\widetilde{\mathbf{s}}$ is $\overline{\mathbf{s}}$; so, having calculated the fuzziness of cluster elaborated with "upper" and "lower" data set the lower and upper values of fuzziness are elaborated in the following way:

$$\overline{\mathbf{s}}_i = \max\left(\widetilde{\mathbf{s}}_i, \underset{\sim}{\mathbf{s}}_i\right) \tag{31}$$

and

$$\underline{\mathbf{s}}_i = \min\left(\widetilde{\mathbf{s}}_i, \underset{\sim}{\mathbf{s}}_i\right). \tag{32}$$

This clustering algorithm creates clusters being type-2 fuzzy sets. The type-2 fuzzy clustering is not widely used. The paper [25] proposes the clustering algorithm. This is modification of the FCM algorithm. The two membership functions are achieved by applying various values of m parameter for criterion function of FCM algorithm. The values of the m parameter are selected by user and are not tuned or modified during the clustering procedure. In our approach the m parameter for both fuzzy sets is constant ($m = 2$), the sets differ by the σ parameter. The gap between "upper" and "lower" fuzzy sets is fitted automatically. The cluster represented by a pair of fuzzy sets can also be regarded as a rough fuzzy set.

3 EXPERIMENTS

The algorithm minimises the criterion function (1). This class of clustering algorithm requires the a priori number of clusters because the algorithm itself cannot elaborate the optimal number of clusters. For experiments the number of clusters is assumed to be 3 or 5.

The f parameter (Equation (1)) has to be fixed a priori. In the experiment we used many values, but the best results were achieved for $f \in [2, 5]$. This interval seems independent from the data set to cluster.

The data tuples with missing values are fully imputed (Section 2.1.2).

3.1 Data Sets

Data sets used in experiments are both synthetic and real life data sets. The synthetic data set is used to check whether the clusters in subspaces are identified correctly. The real life data sets depicting methane concentration (Section 3.1.3) and concrete compressive strength (Section 3.1.2) were prepared and 5 % of values were removed at random. The water treatment plant data set is untouched.

3.1.1 Synthetic Data set

The synthetic data set 'g136' has 6 attributes (dimensions) and two clusters. The first cluster has points generated with Gauss distribution (mean $m = 1$ and standard deviation $\sigma = 1$) in dimenstions 1, 3 and 6. The second cluster (in subspace created by dimensions 2, 4 and 6) is generated with Gauss distribution ($m = 9$, $\sigma = 1$). The not used attribute values are filled with uniform distribution from interval $[0, 10]$. The data set is not normalised.

3.1.2 Concrete Compressive Strength

The data set contains 1050 data tuples describing the parameters of the concrete sample [52]. Each tuple has 9 attributes:

1. the cement content (in [kg/m^3]),
2. blast furnace slag,
3. fly ash,
4. water,
5. superplasticizer,
6. coarse aggregate,
7. fine aggregate,
8. age of sample (in days: 1 – 365) and
9. concrete compressive strength (in [MPa]).

The data set lacks 10 % of values. The data set can be downloaded freely from public repository[1] [13]. Five percent of values are removed from the data sets.

[1] `http://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength`

3.1.3 Methane Concentration

The dataset contains the real life measurements of air parameters in a coal mine in Upper Silesia (Poland). The parameters (measured in 10 second intervals) are:

1. AN31 – the flow of air in the shaft,
2. AN32 – the flow of air in the adjacent shaft,
3. MM32 – concentration of methane (CH_4),
4. production of coal.

To the tuples the 10-minute sums of measurements of (5) AN31, (6) AN32, (7) MM32 are added as dynamic attributes [40]. The data set contains 499 tuples. Five percent of values are removed from the data sets.

3.1.4 Water Treatment Plant Data Set

The data set describes the measurements of sensors in an urban waste water treatment plant. This domain has been stated as an ill-structured domain. Some values are missing from the data set. The attribute values are numeric and continuous. The first attribute (the day of the data) is rejected so there are 38 attributes in 527 tuples. The data set can be downloaded freely from public repository[2] [13]. For brevity of the paper the attributes names are not listed in this paper, but can be found in data description in the above-mentioned repository. The data set originally lacked values.

3.2 Results

Figures 5 and 6 present the results of clustering of synthetic data set 'g136'. It should be stated clearly that the representation in the figures is symbolical. It means two features, namely

1. membership μ of the data tuple and
2. weight z are shown in a combined way.

The figures present the product $\mu \cdot z$ instead of separate figures of μ and z. This approach is only used for better representation of two features in one figure. The attribute weight has no influence on the value of attribute membership. This remark should always be taken into consideration when analysing the above-mentioned figures.

Each rough fuzzy cluster is represented with two Gaussian functions: the wider representing the 'upper' fuzzy set and the narrower for 'lower' fuzzy set.

[2] `http://archive.ics.uci.edu/ml/datasets/Water+Treatment+Plant`

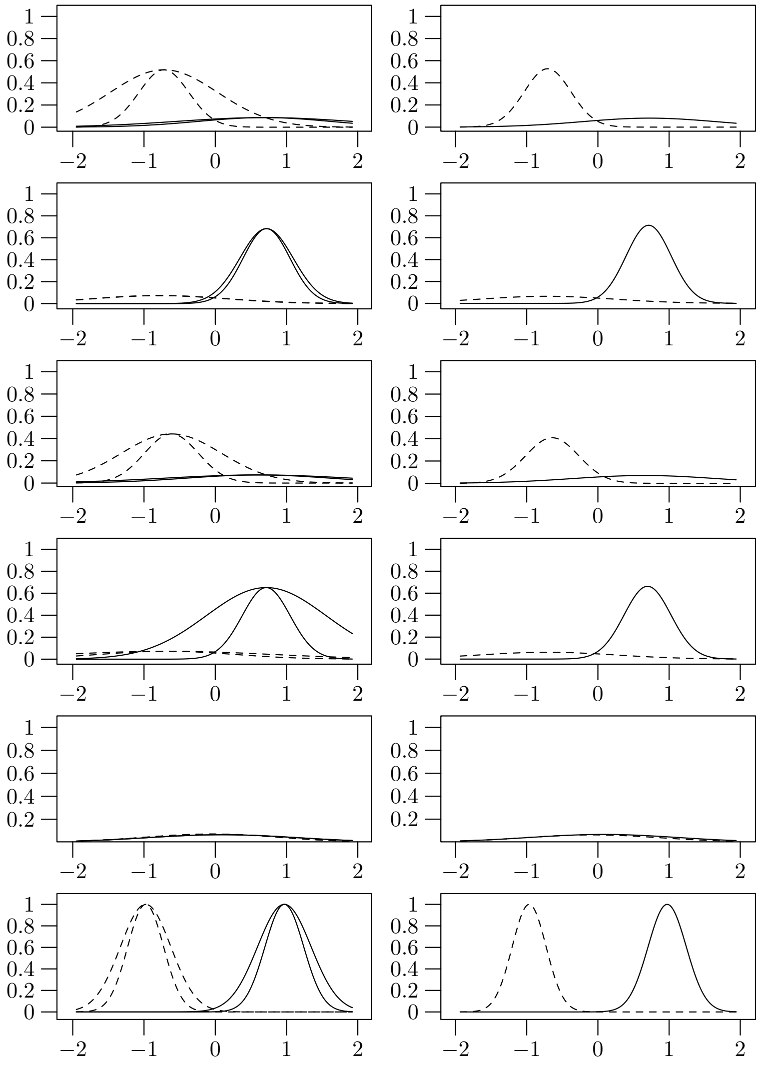

Figure 5. The results of clustering of synthetic data set 'g136' into two clusters. The left column presents clustering of 'g136' data set with 10 % of values missing from attributes 1, 3 and 4. The attributes 2, 5 and 6 miss no values. The right column presents the results of clustering of 'g136' data set with no missing values. The representation of cluster membership functions is symbolical – the membership function is combined with the data tuple's weight (see details at the beginning of Section 3.2).

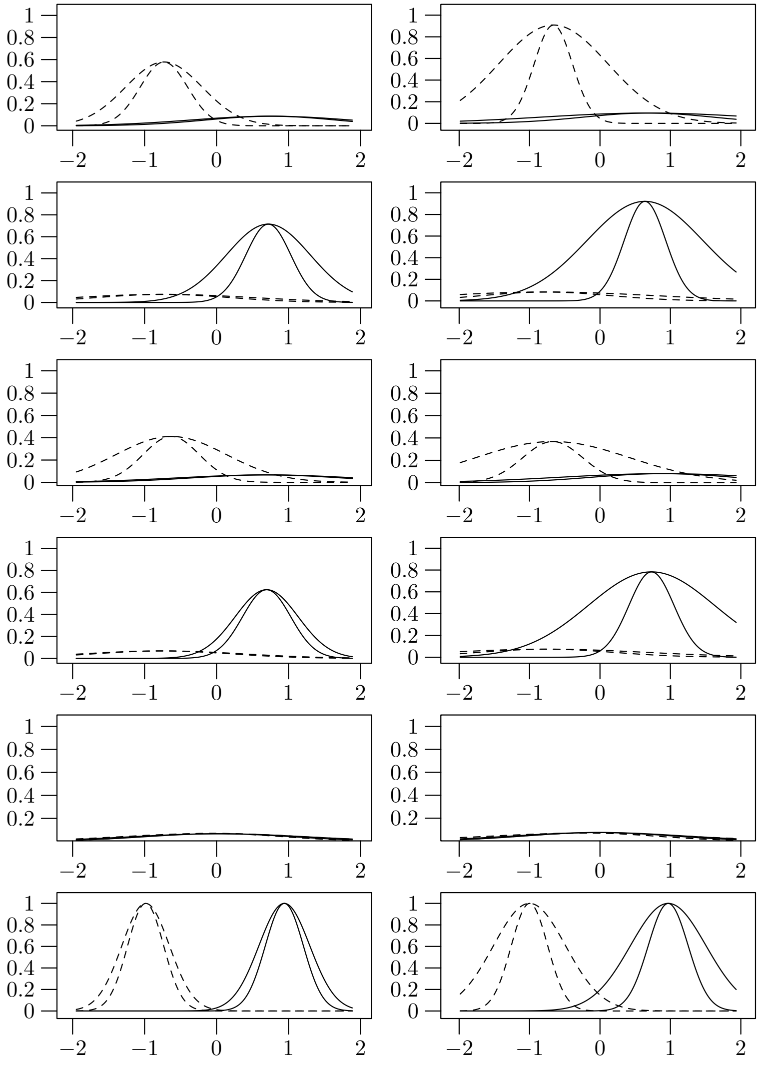

Figure 6. The results of clustering of synthetic data set 'g136' into two clusters. The left column presents clustering of 'g136' data set with 5 % of values missing from all attributes. The right column presents the results of clustering of 'g136' data set with 10 % of values missing from all attributes. The representation of cluster membership functions is symbolical – the membership function is combined with the data tuple's weight (see details at the beginning of the Section 3.2).

The results of rough fuzzy clustering should be analysed in two aspects. The first one is the proper identification of subspaces. The results presented in Figures 5 and 6 reveal correct identification of subspaces in dimensions: 1-3-6 and 2-4-6. The 5^{th} dimension is not used correctly in either cluster. The second aspect is correct handling of missing values. The algorithm creates rough fuzzy sets. The results show that the more values are missing, the wider is the separation of 'upper' and 'lower' data sets. If all data tuples are complete and lack no values, the rough fuzzy clustering degenerates to fuzzy clustering (see the right column of Figure 5). However, if there are attributes with all values and attributes with missing values in one data set, the rough fuzzy sets are elaborated for both kinds of attributes. This feature can be observed in the left column of Figure 5, where attributes 2, 5 and 6 lack no values, but are represented by the rough fuzzy sets. It can also be observed that the roughness of sets for the complete attributes is less than roughness of attributes with missing values (attributes 1, 3 and 4).

The results of clustering of real life data sets introduced in Section 3.1 are presented in Tables 2, 3, 4 and 5.

The cluster parameters (cluster centre (v), "lower" ($\underline{s}$) and "upper" ($\overline{s}$) fuzziness and attribute's weight (z)) are presented in symbolic way as $v_{\underline{s}}^{\overline{s}}(z)$.

The f parameter in clustering algorithm influences the weight values. If $f = 1$ only one attribute in the cluster is assigned with maximal (1) value and all other attributes get zero weights. This is explained by Equation (26). If $f \gg 1$ (in practice $f > 10$) then all attributes get similar weights (after augmenting all weights are almost equal to 1). This is confirmed by multiple experiments.

	clusters				
attr.	I	II	III	IV	V
1	$-0.16_{0.39}^{0.89}(0.90)$	$-1.31_{0.29}^{0.58}(0.92)$	$-0.84_{0.39}^{0.90}(0.95)$	$0.27_{0.48}^{1.11}(0.88)$	$1.25_{0.27}^{0.52}(0.88)$
2	$0.15_{0.39}^{0.79}(0.89)$	$1.37_{0.33}^{0.65}(0.85)$	$0.82_{0.39}^{0.81}(0.95)$	$-0.31_{0.50}^{1.00}(0.85)$	$-1.20_{0.32}^{0.56}(0.81)$
3	$0.10_{0.49}^{0.83}(0.79)$	$1.37_{0.68}^{1.15}(0.60)$	$0.74_{0.54}^{0.91}(0.81)$	$-0.61_{0.40}^{0.70}(0.95)$	$-0.92_{0.53}^{0.84}(0.63)$
4	$0.10_{0.84}^{1.15}(0.61)$	$0.58_{0.65}^{0.90}(0.61)$	$0.04_{0.86}^{1.19}(0.64)$	$-0.70_{0.88}^{1.27}(0.65)$	$0.03_{1.12}^{1.40}(0.43)$
5	$-0.18_{0.34}^{0.86}(0.96)$	$-1.35_{0.24}^{0.53}(1.00)$	$-0.88_{0.35}^{0.87}(1.00)$	$0.28_{0.39}^{1.03}(0.97)$	$1.27_{0.21}^{0.44}(1.00)$
6	$0.15_{0.31}^{0.78}(1.00)$	$1.44_{0.31}^{0.68}(0.88)$	$0.85_{0.35}^{0.86}(1.00)$	$-0.33_{0.37}^{0.98}(1.00)$	$-1.24_{0.21}^{0.43}(1.00)$
7	$0.10_{0.48}^{0.61}(0.80)$	$1.41_{0.67}^{0.78}(0.60)$	$0.77_{0.53}^{0.69}(0.82)$	$-0.60_{0.41}^{0.60}(0.95)$	$-0.94_{0.52}^{0.56}(0.64)$

Table 2. Results for 'methane' data set, $f = 5$

The 'methane' data set was clustered into 5 clusters, the results are presented in Table 2 ($f = 5$). The data set lacks 5 % of values. The very important phenomenon can be seen in the results of clustering (Table 2). The most significant attribute in four clusters is the 5^{th} one (10-minute sum of the flow of air in the shaft AN31). This is correlated with the high importance of the first attribute – the flow of the air in the shaft AN31. In one cluster the most important attribute is the 3rd one (MM32 – concentration of methane) and the second most important one is the 7^{th} attribute (10-minute sums of measurements of MM32). In the second cluster the similar correlation between the 6^{th} (sum of AN32) and 2nd (AN32 – the flow of air in the adjacent shaft) attribute can be easily noticed. In is worth mentioning that in 4 clusters the attribute describing sum of measurements is more important that the attribute expressing the measurements themselves.

	clusters				
attr.	I	II	III	IV	V
1	$-0.54^{1.08}_{0.80}(0.00)$	$-0.45^{1.10}_{0.43}(0.01)$	$0.28^{1.25}_{1.06}(0.00)$	$0.33^{1.29}_{1.13}(0.00)$	$-0.67^{1.31}_{0.68}(0.15)$
2	$-0.04^{1.22}_{0.86}(0.00)$	$-0.85^{0.13}_{0.04}(1.00)$	$0.31^{1.47}_{1.11}(0.00)$	$-0.41^{1.08}_{0.34}(0.03)$	$-0.03^{1.36}_{0.37}(0.53)$
3	$1.10^{1.45}_{0.53}(0.00)$	$1.01^{1.25}_{0.40}(0.01)$	$-0.84^{0.04}_{0.00}(1.00)$	$0.98^{1.70}_{0.58}(0.01)$	$1.03^{1.75}_{0.83}(0.10)$
4	$0.09^{1.42}_{0.98}(0.00)$	$-0.58^{1.08}_{0.70}(0.00)$	$0.26^{1.24}_{0.99}(0.00)$	$-0.63^{1.27}_{0.92}(0.00)$	$-0.81^{1.18}_{0.66}(0.16)$
5	$0.56^{1.12}_{0.56}(0.00)$	$0.38^{1.17}_{0.53}(0.01)$	$-0.35^{1.40}_{1.14}(0.00)$	$0.45^{1.27}_{0.57}(0.01)$	$0.39^{1.08}_{0.38}(0.50)$
6	$-0.52^{1.43}_{1.01}(0.00)$	$0.52^{1.01}_{0.60}(0.00)$	$-0.03^{1.36}_{0.96}(0.00)$	$-0.13^{1.34}_{0.99}(0.00)$	$0.62^{1.02}_{0.71}(0.14)$
7	$-0.18^{1.25}_{0.87}(0.00)$	$0.69^{0.96}_{0.70}(0.00)$	$-0.09^{1.41}_{1.09}(0.00)$	$0.01^{1.42}_{0.58}(0.01)$	$0.02^{0.84}_{0.27}(1.00)$
8	$-0.27^{0.08}_{0.02}(1.00)$	$-0.01^{1.38}_{0.61}(0.00)$	$0.11^{1.56}_{1.22}(0.00)$	$-0.63^{0.16}_{0.06}(1.00)$	$0.40^{1.67}_{0.38}(0.50)$
9	$-0.15^{0.80}_{0.78}(0.00)$	$-0.11^{0.89}_{0.82}(0.00)$	$0.01^{1.13}_{1.10}(0.00)$	$-0.51^{0.66}_{0.62}(0.01)$	$1.02^{0.91}_{0.58}(0.21)$

Table 3. Results for 'concrete' data set, $f = 2$

	clusters				
attr.	I	II	III	IV	V
1	$-0.53^{1.00}_{0.81}(0.01)$	$-0.84^{1.06}_{0.33}(0.85)$	$0.28^{1.19}_{1.06}(0.00)$	$-0.48^{0.91}_{0.34}(0.02)$	$1.07^{1.19}_{0.83}(0.25)$
2	$-0.00^{1.11}_{0.87}(0.01)$	$0.02^{1.51}_{0.35}(0.82)$	$0.31^{1.35}_{1.11}(0.00)$	$-0.86^{0.00}_{0.00}(1.00)$	$-0.61^{0.19}_{0.05}(1.00)$
3	$1.10^{1.24}_{0.53}(0.01)$	$0.98^{1.61}_{0.84}(0.53)$	$-0.84^{0.00}_{0.00}(1.00)$	$1.01^{1.09}_{0.39}(0.02)$	$0.97^{1.00}_{0.42}(0.35)$
4	$0.12^{1.31}_{0.98}(0.01)$	$-0.82^{1.15}_{0.75}(0.56)$	$0.26^{1.16}_{0.99}(0.00)$	$-0.59^{1.00}_{0.68}(0.02)$	$-0.44^{1.05}_{0.89}(0.24)$
5	$0.56^{0.98}_{0.57}(0.01)$	$0.37^{1.06}_{0.38}(0.79)$	$-0.36^{1.32}_{1.14}(0.00)$	$0.39^{1.05}_{0.52}(0.02)$	$0.46^{1.10}_{0.63}(0.29)$
6	$-0.57^{1.32}_{1.00}(0.01)$	$0.82^{1.04}_{0.43}(0.74)$	$-0.03^{1.24}_{0.96}(0.00)$	$0.56^{0.88}_{0.55}(0.02)$	$-0.79^{0.82}_{0.76}(0.26)$
7	$-0.22^{1.15}_{0.86}(0.01)$	$0.04^{0.99}_{0.24}(1.00)$	$-0.09^{1.31}_{1.09}(0.00)$	$0.71^{0.87}_{0.69}(0.02)$	$-0.08^{0.94}_{0.70}(0.27)$
8	$-0.27^{0.00}_{0.00}(1.00)$	$-0.13^{1.32}_{0.56}(0.65)$	$0.11^{1.45}_{1.22}(0.00)$	$-0.05^{1.13}_{0.62}(0.02)$	$-0.38^{0.79}_{0.40}(0.37)$
9	$-0.14^{0.79}_{0.79}(0.01)$	$0.02^{1.02}_{0.97}(0.48)$	$0.00^{1.12}_{1.10}(0.00)$	$-0.17^{0.89}_{0.83}(0.01)$	$0.22^{0.91}_{0.90}(0.24)$

Table 4. Results for 'concrete' data set, $f = 5$

The clusters elaborated for 'concrete' data set are presented in Table 3 for $f = 2$ and Table 4 for $f = 5$. For both values of f the most important attributes are 2 (blast furnace slag), 3 (fly ash), 7 (fine aggregate) and 8 (age of sample). Clustering with $f = 5$ gives higher values of attributes weights. This is also clearly seen in Table 3 where some singular attributes achieve maximal weights whereas the rest have very low weights. When $f = 5$ (Table 4) the weights of attributes are significantly higher (with exception of cluster III). This is also important, because when the attributes have higher weight the reliability of cluster's parameters for this attribute is higher.

The results of clustering for the 'water' data set are presented in Table 5. The attributes with weights $\geqslant 0.75$ are printed in negative for easier identification. The results of clustering into 3 clusters show that in the 2^{nd} and 3^{rd} cluster the most important attribute is 'output sediments' (No. 28), whereas in the 1^{st} cluster this attribute has not very high importance. The importance of attributes in various clusters is (to some extend) complementary. In the 1st cluster the weights assigned to attributes are relatively high, whereas the weights in clusters 2 and 3 are lower. The precision of localisation of cluster centre, whose weight is low, is not high. The higher the weight, the more reliable the localisation and fuzziness of cluster.

The clustering algorithm is parametrized. The f parameter seems to be the most important. The experiments show that the best interval for this parameter is $[2, 5]$. The values less than 2 lead to selection of only one attribute. The values greater than 5 assign similar weights to all attributes; this reduces the rough fuzzy subspace clustering to rough fuzzy clustering without subspace identification. If the experimentator wishes sharper clustering, the smaller values of f should be used. If the subspaces should not be identified very sharply the higher values of f should be used. Cluster identification sharpness can be different in one clustering experiment. This phenomenon can be observed in Table 4. The subspace for the 1st cluster is identified sharply, whereas the 5th cluster in the same experiment has far less sharp subspace. Based on the results of clustering, it seems that the optimal interval $[2, 5]$ for parameter f is independent from the clustered data set.

4 CONCLUSIONS

The paper presents rough fuzzy subspace clustering algorithm with experimental results. To handle the lack of attribute values three approaches are used: marginalisation, imputation and rough sets. The algorithm also assigns weights to attributes in each cluster; this leads to subspace clustering. The parameters of clusters are elaborated in the iterative procedure based on minimising the criterion function. The number of clusters has to be given a priori. The crucial parameter of the proposed algorithm is the parameter having the influence on the sharpness of elaborated subspace cluster. The lower values of the parameter lead to selection of the most important attribute. The higher values create clustering in global space, not in subspaces. This parameter seems to be independent from the data sets to cluster.

Clustering of synthetic data sets reveals the ability of the proposed algorithm to identify the subspaces. Assigning weights to the attributes can also find the relations between the attributes. Besides, the algorithm is capable of handling the data with missing values. The more values miss from the data set the more rough clusters are elaborated. The fact of missing some values from one attribute may interfere the complete attributes with no missing values. In such data sets (with complete attributes and attributes with missing values) the attributes with no missing values may be modelled with rough sets.

The algorithm is thought to be one of the clustering algorithms used for input domain partition in neuro-fuzzy systems.

Acknowledgements

The author is grateful to the anonymous referees for their constructive comments that have helped to improve the paper.

	clusters		
attr.	I	II	III
1	$-0.10^{1.06}_{0.87}(0.74)$	$0.08^{1.09}_{0.98}(0.27)$	$0.02^{1.04}_{1.00}(0.30)$
2	$-0.12^{0.82}_{0.73}(0.81)$	$0.01^{0.86}_{0.78}(0.31)$	$-0.02^{0.88}_{0.79}(0.33)$
3	$-0.12^{0.86}_{0.80}(0.77)$	$0.07^{0.96}_{0.95}(0.29)$	$0.15^{1.01}_{0.98}(0.31)$
4	$0.22^{2.54}_{0.93}(0.72)$	$-0.03^{1.96}_{0.99}(0.29)$	$0.01^{2.09}_{0.98}(0.31)$
5	$0.08^{1.13}_{0.99}(0.70)$	$-0.01^{1.14}_{1.03}(0.28)$	$-0.03^{1.12}_{0.95}(0.32)$
6	$-0.14^{0.54}_{0.48}(1.00)$	$0.35^{1.45}_{1.31}(0.24)$	$-0.11^{0.83}_{0.60}(0.40)$
7	$0.17^{0.96}_{0.68}(0.84)$	$-0.34^{1.68}_{1.23}(0.26)$	$0.06^{1.39}_{0.97}(0.31)$
8	$-0.15^{0.62}_{0.54}(0.88)$	$0.42^{1.90}_{1.26}(0.21)$	$-0.06^{0.65}_{0.64}(0.38)$
9	$-0.32^{1.12}_{0.69}(0.83)$	$0.01^{1.17}_{1.03}(0.28)$	$0.10^{1.15}_{1.06}(0.30)$
10	$-0.10^{0.84}_{0.79}(0.76)$	$0.09^{0.96}_{0.85}(0.29)$	$0.19^{1.04}_{0.88}(0.30)$
11	$0.18^{2.62}_{0.94}(0.71)$	$0.03^{2.05}_{0.96}(0.29)$	$0.05^{2.19}_{0.96}(0.31)$
12	$-0.17^{0.52}_{0.37}(0.96)$	$0.42^{1.70}_{1.06}(0.22)$	$-0.09^{0.68}_{0.48}(0.37)$
13	$0.20^{1.07}_{0.73}(0.81)$	$-0.37^{1.67}_{1.21}(0.26)$	$0.04^{1.43}_{0.98}(0.31)$
14	$-0.07^{0.59}_{0.52}(0.90)$	$0.41^{1.91}_{1.16}(0.21)$	$-0.09^{0.67}_{0.61}(0.38)$
15	$-0.31^{1.15}_{0.70}(0.83)$	$-0.00^{1.20}_{1.02}(0.28)$	$0.10^{1.19}_{1.06}(0.30)$
16	$-0.21^{0.86}_{0.76}(0.75)$	$0.16^{0.94}_{0.86}(0.29)$	$0.22^{1.03}_{0.88}(0.30)$
17	$0.34^{2.50}_{0.94}(0.72)$	$-0.11^{1.99}_{0.98}(0.29)$	$0.02^{2.13}_{0.98}(0.31)$
18	$0.27^{1.24}_{0.96}(0.71)$	$-0.15^{1.29}_{0.95}(0.29)$	$0.02^{1.32}_{0.95}(0.32)$
19	$0.09^{0.94}_{0.87}(0.74)$	$0.03^{1.28}_{1.06}(0.28)$	$-0.04^{1.21}_{0.95}(0.32)$
20	$0.12^{1.21}_{0.70}(0.83)$	$-0.21^{1.81}_{1.13}(0.27)$	$0.04^{1.65}_{0.97}(0.31)$
21	$0.10^{1.19}_{0.93}(0.72)$	$-0.05^{1.11}_{0.93}(0.30)$	$0.02^{1.23}_{1.06}(0.30)$
22	$-0.28^{1.06}_{0.72}(0.82)$	$-0.00^{1.15}_{1.03}(0.28)$	$0.09^{1.14}_{1.06}(0.30)$
23	$-0.31^{0.96}_{0.77}(0.79)$	$0.19^{1.17}_{0.80}(0.32)$	$0.11^{1.19}_{0.81}(0.34)$
24	$0.51^{3.79}_{1.38}(0.59)$	$-0.17^{1.84}_{0.31}(0.51)$	$-0.13^{2.00}_{0.33}(0.54)$
25	$0.64^{2.02}_{1.55}(0.56)$	$-0.25^{1.33}_{0.66}(0.35)$	$-0.14^{1.35}_{0.71}(0.37)$
26	$0.88^{2.42}_{1.53}(0.56)$	$-0.25^{0.96}_{0.40}(0.45)$	$-0.19^{0.94}_{0.43}(0.47)$
27	$-0.09^{1.26}_{0.72}(0.82)$	$-0.12^{1.84}_{1.09}(0.27)$	$-0.00^{1.71}_{1.03}(0.30)$
28	$1.11^{3.34}_{1.95}(0.38)$	$-0.12^{0.11}_{0.08}(1.00)$	$-0.11^{0.14}_{0.09}(1.00)$
29	$-0.27^{1.10}_{0.68}(0.84)$	$-0.02^{1.14}_{1.04}(0.28)$	$0.08^{1.14}_{1.05}(0.30)$
30	$-0.23^{2.62}_{1.07}(0.67)$	$0.13^{2.14}_{0.97}(0.29)$	$-0.01^{2.27}_{0.99}(0.31)$
31	$-0.31^{1.11}_{0.77}(0.66)$	$0.33^{1.21}_{1.06}(0.28)$	$-0.04^{1.08}_{0.95}(0.32)$
32	$-0.13^{1.06}_{0.78}(0.67)$	$0.15^{0.87}_{0.84}(0.31)$	$-0.05^{1.06}_{0.91}(0.30)$
33	$-0.55^{2.93}_{1.60}(0.55)$	$0.17^{2.03}_{0.57}(0.38)$	$0.17^{2.20}_{0.60}(0.40)$
34	$-0.48^{1.57}_{1.30}(0.61)$	$0.17^{1.40}_{0.77}(0.33)$	$0.16^{1.38}_{0.81}(0.34)$
35	$-0.57^{2.88}_{1.84}(0.51)$	$0.19^{1.97}_{0.50}(0.40)$	$0.16^{2.13}_{0.52}(0.43)$
36	$-0.63^{1.52}_{1.49}(0.56)$	$0.23^{1.37}_{0.76}(0.33)$	$0.12^{1.34}_{0.79}(0.35)$
37	$-0.88^{1.85}_{1.41}(0.58)$	$0.34^{1.09}_{0.49}(0.41)$	$0.16^{1.03}_{0.51}(0.43)$
38	$-1.13^{2.97}_{2.27}(0.40)$	$0.14^{0.11}_{0.09}(0.98)$	$0.12^{0.14}_{0.12}(0.89)$

Table 5. Results for the 'water' data set ($f = 5$). The attributes with weights $\geqslant 0.75$ are printed white in black boxes.

REFERENCES

[1] Acuña, E.—Rodriguez, C.: The Treatment of Missing Values and its Effect in the Classifier Accuracy. In D. Banks, L. House, F. R. McMorris, P. Arabie, W. Gaul (Eds.): Classification, Clustering and Data Mining Applications, Springer 2004, pp. 639–648.

[2] Aggarwal, C. C.—Wolf, J. L.—Yu, P. S.—Procopiuc, C.—Park, J. S.: Fast Algorithms for Projected Clustering. SIGMOD Rec., Vol. 28, 1999, No. 2, pp. 61–72.

[3] Aggarwal, C. C.—Yu, P. S.: Finding Generalized Projected Clusters in High Dimensional Spaces. In SIGMOD '00: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, New York, NY, USA 2000, pp. 70–81.

[4] Agrawal, R.—Gehrke, J.—Gunopulos, D.—Raghavan, P.: Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications. SIGMOD Rec., Vol. 27, 1998, No. 2, pp. 94–105.

[5] Batista, G. E. A.—Monard, M. C.: An Analysis of Four Missing Data Treatment Methods for Supervised Learning. Applied Artificial Intelligence, Vol. 17, 2003, No. 5-6, pp. 519–533.

[6] Chan, L. S.—Gilman, J. A.—Dunn, O. J.: Alternative Approaches to Missing Values in Discriminant Analysis. Journal of the American Statistical Association, Vol. 71, 1976, No. 356, pp. 842–844.

[7] Cheng, C. H.—Fu, A.-W.—Zhang, Y.: Entropy-Based Subspace Clustering for Mining Numerical Data. In KDD '99: Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York 1999, pp. 84–93.

[8] Cooke, M.—Green, P.—Josifovski, L.—Vizinho, A.: Robust Automatic Speech Recognition with Missing and Unreliable Acoustic Data. Speech Communication, Vol. 34, 2001, pp. 267–285.

[9] Czogała, E.—Łeski, J.: Fuzzy and Neuro-Fuzzy Intelligent Systems. Series in Fuzziness and Soft Computing, Physica-Verlag, A Springer-Verlag Company, Heidelberg, New York 2000.

[10] Dempster, A. P.—Laird, N. M.—Rubin, D. B.: Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society, Series B, Vol. 39, 1977, No. 1, pp. 1–38.

[11] Farhangfar, A.—Kurgan, L. A.—Pedrycz, W.: A Novel Framework for Imputation of Missing Values in Databases. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, Vol. 37, 2007, No. 5, pp. 692–709.

[12] Farhangfar, A.—Kurgan, L.—Dy, J.: Impact of Imputation of Missing Values on Classification Error for Discrete Data. Pattern Recognition, Vol. 41, 2008, No. 12, pp. 3692–3705.

[13] Frank, A.—Asuncion, A.: UCI Machine Learning Repository, 2010.

[14] Friedman, J. H.—Meulman, J. J.: Clustering Objects on Subsets of Attributes. Journal of the Royal Statistical Society: Series B, Vol. 66, 2004, pp. 815–849.

[15] Fuller, W. A.—Kim, J. K.: Hot Deck Imputation for the Response Model. Survey Methodology, Vol. 31, 2005, No. 2, pp. 139–149.

[16] Gan, G.—Wu, J.: A Convergence Theorem for the Fuzzy Subspace Clustering (FSC) Algorithm. Pattern Recogn., Vol. 41, 2008, No. 6, pp. 1939–1947.

[17] Gan, G.—Wu, J.—Yang, Z.: A Fuzzy Subspace Algorithm for Clustering High Dimensional Data. In Proceedings of Second International Conference on Advanced Data Mining and Applications (ADMA) 2006, China 2006, Vol. 4093 of Lecture Notes in Computer Science, Springer 2006, pp. 271–278.

[18] Ghahramani, Z.—Jordan, M. I.: Learning from Incomplete Data. Technical report, Lab Memo No. 1509, CBCL Paper No. 108, MIT AI Lab 1995.

[19] Goil, S.—Nagesh, H.—Choudhary, A.: Mafia: Efficient and Scalable Subspace Clustering for Very Large Data Sets. Technical report 1999.

[20] Grzymała-Busse, J.: A Rough set Approach to Data With Missing Attribute Values. In Guoying Wang, James Peters, Andrzej Skowron, and Yiyu Yao (Eds.): Rough Sets and Knowledge Technology, Vol. 4062 of Lecture Notes in Computer Science, Springer 2006, pp. 58–67.

[21] Grzymała-Busse, J.—Goodwin, L.—Grzymala-Busse, W.—Zheng, X.: Handling Missing Attribute Values in Preterm Birth Data Sets. In Dominik Slezak, JingTao Yao, James Peters, Wojciech Ziarko, and Xiaohua Hu (Eds.): Rough Sets, Rough fuzzy subspace clustering for data with missing values, Vol. 3642 of Lecture Notes in Computer Science, Springer 2005, pp. 342–351.

[22] Grzymała-Busse, J.—Hu, M.: A Comparison of Several Approaches to Missing Attribute Values in Data Mining. In Wojciech Ziarko and Yiyu Yao (Eds.): Rough Sets and Current Trends in Computing, Vol. 2005 of Lecture Notes in Computer Science, Springer 2001, pp. 378–385.

[23] Hathaway, R. J.—Bezdek, J. C.: Fuzzy C-Means Clustering of Incomplete Data. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 31, 2001, No. 5, pp. 735–744.

[24] Himmelspach, L.—Conrad, S.: Fuzzy Clustering of Incomplete Data Based on Cluster Dispersion. In Eyke Hüllermeier, Rudolf Kruse, and Frank Hoffmann (Eds.): Computational Intelligence for Knowledge-Based Systems Design, Vol. 6178 of Lecture Notes in Computer Science, Springer 2010, pp. 59–68.

[25] Hwang, C.—Chung-Hoon Rhee, F.: An Interval Type-2 Fuzzy c Spherical Shells Algorithm. In Proceedings of the 2004 IEEE International Conference on Fuzzy Systems, Budapest, Hungary 2004, pp. 1117–1122.

[26] Jang, J. S. R.: ANFIS: Adaptive-Network-Based Fuzzy Inference System. IEEE Transactions on Systems, Man, and Cybernetics, Vol. 23, 1993, No. 3, pp. 665–684.

[27] Kalton, G.—Kasprzyk, D.: The Treatment of Missing Survey Data. Survey Methodology, Vol. 12, 1986, pp. 1–16.

[28] Lakshminarayan, K.—Harp, S. A.—Samad, T.: Imputation of Missing Data in Industrial Databases. Applied Intelligence, Vol. 11, 1999, No. 3, pp. 259–275, 10.1023/A:1008334909089.

[29] Łeski, J.: Neuro-Fuzzy Systems. Wydawnictwa Naukowo-Techniczne, Warszawa, 2008 (in Polish).

[30] Little, R. J.—Rubin, D. B.: Statistical Analysis with Missing Data. John Wiley and Sons, New York 1987.

[31] Luengo, J.—García, S.—Herrera, F.: A Study on the use of Imputation Methods for Experimentation With Radial Basis Function Network Classifiers Handling Missing Attribute Values: The Good Synergy Between Rbfns and Eventcovering Method. Neural Networks, Vol. 23, 2010, No. 3, pp. 406–418.

[32] Luengo, J.—García, S.—Herrera, F.: On the Choice of the Best Imputation Methods for Missing Values Considering Three Groups of Classification Methods. Knowledge and Information Systems 2011, pp. 1–32.

[33] Mundfrom, D. J.—Whitcomb, A.: Imputing Missing Values: The Effect on the Accuracy of Classification. Multiple Linear Regression Viewpoints, Vol. 25, 1998, No. 1, pp. 13–19.

[34] Nelles, O.—Fink, A.—Babuška, R.—Setnes, M.: Comparison of Two Construction Algorithms for Takagi-Sugeno Fuzzy Models. International Journal of Applied Mathematics and Computer Science, Vol. 10, 2000, No. 4, pp. 835–855.

[35] Nowicki, R.: Rough-Neuro-Fuzzy System with MICOG Defuzzification. In 2006 IEEE International Conference on Fuzzy Systems, Vancouver, Canada 2006, pp. 1958–1965.

[36] Parsons, L.—Haque, E.—Liu, H.: Subspace Clustering for High Dimensional Data: A Review. SIGKDD Explor. Newsl., Vol. 6, 2004, No. 1, pp. 90–105.

[37] Pedrycz, W.: Conditional Fuzzy Clustering in the Design of Radial Basis Function Neural Networks. IEEE Transactions on Neural Networks, Vol. 9, 1998, No. 4, pp. 601–612.

[38] Renz, C.—Rajapakse, C.—Razvi, K.—Liang, S. K. C.: Ovarian Cancer Classification with Missing Data. In Proceedings of the 9$^{\text{th}}$ International Conference on Neural Information Processing (ICONIP ’02), Singapore 2002, Vol. 2, pp. 809–813.

[39] Rubin, D. B.: Multiple Imputation For Nonresponse In Surveys. John Wiley & Sons 1987.

[40] Sikora, M.—Sikora, B.: Application of Machine Learning for Prediction of Methane Concentration in a Coal-Mine. Archives of Mining Sciences, Vol. 51, 2006, No. 4, pp. 475–492.

[41] Siminski, K.: Neuro-Fuzzy System with Hierarchical Domain Partition. In Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation (CIMCA 2008), Vienna, Austria 2008, pp. 392–397.

[42] Siminski, K.: Clustering in Fuzzy Subspaces. Theoretical and Applied Informatics, Vol. 24, 2012, No. 4, pp. 313–326.

[43] Siminski, K.: Neuro-Rough-Fuzzy Approach for Regression Modelling from Missing Data. International Journal of Applied Mathematics and Computer Science, Vol. 22, 2012, No. 2, pp. 461–476.

[44] Siminski, K.: Clustering with Missing Values. Fundamenta Informaticae, Vol. 123, 2013, No. 3, pp. 331–350.

[45] Siminski, K.: Neuro-Fuzzy System with Weighted Attributes. Soft Computing 2013.

[46] Troyanskaya, O.—Cantor, M.—Sherlock, G.—Brown, P.—Hastie, T.—Tibshirani, R.—Botstein, D.—Altman, R. B.: Missing Value Estimation Methods for DNA Microarrays. Bioinformatics, Vol. 17, 2001, No. 6, pp. 520–525.

[47] Wagstaff, K.: Clustering with Missing Values: No Imputation Required. In D. Banks, L. House, F. R. Mcmorris, P. Arabie, and W. Gaul (Eds.): Classification, Clustering, and Data Mining Applications. Proceedings of the Meeting of the International Federation of Classification Societies, Springer 2004, pp. 649–658.

[48] Wagstaff, K. L.—Laidler, V. G.: Making the Most of Missing Values: Object Clustering with Partial Data in Astronomy. In Proceedings of Astronomical Data Analysis Software and Systems XIV, Pasadena, California, USA 2005, Vol. 347, pp. 172–176.

[49] Wong, A. K. C.—Chiu, D. K. Y.: Synthesizing Statistical Knowledge from Incomplete Mixed-Mode Data. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 9, 1987, pp. 796–805.

[50] Wuwongse, V.—Veluppilai, S.: Fuzzy Modeling with Genetic Algorithms. Computers and Artificial Intelligence, Vol. 16, 1997, No. 3, pp. 275–289.

[51] Yang, J.—Wang, W.—Wang, X.—Yu, P.: δ-Clusters: Capturing Subspace Correlation in a Large Data Set. In, 2002. Proceedings of 18$^{\text{th}}$ International Conference on, Data Engineering 2002, pp. 517–528.

[52] Yeh, I. C.: Modeling of Strength of High-Performance Concrete Using Artificial Neural Networks. Cement and Concrete Research, Vol. 28, 1998, No. 12, pp. 1797–1808.

[53] Zhang, C.—Zhu, X.—Zhang, J.—Qin, Y.—Zhang, S.: GBKII: An Imputation Method for Missing Values. Advances in Knowledge Discovery and Data Mining, Vol. 4426, 2007, pp. 1080–1087.

[54] Zhang, S.: Shell-Neighbor Method and Its Application in Missing Data Imputation. Applied Intelligence, Vol. 35, 2011, No. 1, pp. 1–11, 10.1007/s10489-009-0207-6.

Krzysztof Siminski received the M. Sc. and Ph. D. degrees in computer science from the Silesian University of Technology (Gliwice, Poland) in 2006 and 2009, respectively. His main interests include data mining, fuzzy reasoning, natural language processing.