Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Fuzzy Sets and Systems ••• (••••) •••–•••
FUZZY sets and systems
www.elsevier.com/locate/fss

# Fuzzy weighted C-ordered means clustering algorithm

Krzysztof Siminski

*Institute of Informatics, Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, Poland*

Received 21 April 2016; received in revised form 11 November 2016; accepted 7 January 2017

**Abstract**

In real life data sets some attributes may have lower importance or even may be completely noninformative. The subspace clustering algorithms have been proposed to handle this. The soft subspace algorithms are vulnerable to noise and outliers. The paper presents a novel algorithm that handles both various importance of attributes and outliers. The proposed Fuzzy Weighted C-Ordered Mean (FWCOM) clustering algorithm elaborates clusters in soft subspaces. In each cluster each attribute is assigned a weight from interval [0, 1]. Each attribute has its individual weight (importance) in each cluster. The proposed algorithm applies the ordering technique to effectively reduce the influence of outliers and noise. The paper is accompanied by numerical experiments.


*Keywords:* Fuzzy clustering; Subspace clustering; Ordered weighted averaging

## 1. Introduction

The real life data are far from being ideal. Data granules may be embedded in some subspaces of the task space. Some of dimensions may have lower importance or may even be totally noninformative and superfluous. The global reduction of task dimensionality by feature transformation (e.g. Principal Component Analysis or Singular Value Decomposition) may lead to problems with interpretation of elaborated results. The global approach may not be satisfactory because noninformative dimensions in one data granule may be of high importance in the other one. This leads to subspace clustering [7,10,14] for elaboration of data granules in subspaces of the original task space.

Many subspace algorithms can be classified as top-down or bottom-up techniques. The top-down algorithms start with all dimensions and try to throw away dimensions of lower importance (e.g. PROCLUS [1], ORCLUS [2], $\delta$-Clusters [22]. The bottom-up approach splits the data with a grid, tests the density of regions, and extract relevant dimensions (e.g. CLIQUE [3], ENCLUS [4], MAFIA [11]). In algorithms mentioned above the weights of dimensions in clusters is either 0 or 1 (hard weights).

The crisp (hard) weights of attributes may not be satisfactory in some applications. Algorithms that elaborate fuzzy (soft, nonbinary) weights for attributes are an interesting direction of research. Some algorithm have been recently

*E-mail address:* Krzysztof.Siminski@polsl.pl.

http://dx.doi.org/10.1016/j.fss.2017.01.001


Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001

Table 1
Symbols used in the paper.

| Symbol | Meaning |
|---|---|
| $\mathbb{C}$ | set of clusters |
| $C$ | number of clusters, $C = \\|\mathbb{C}\\|$ |
| $\mathbb{X}$ | set of data items, $\mathbb{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_X\}$ |
| $X$ | number of data items, $X = \\|\mathbb{X}\\|$ |
| $\mathbf{x}$ | data item, $\mathbf{x} = [x_1, \ldots, x_D]$ |
| $D$ | number of attributes |
| $\mathbf{U}$ | membership matrix $[C \times X]$ |
| $u_{cx}$ | membership of the $x$th item to the $c$th cluster |
| $\mathbf{V}$ | matrix of cluster centres, $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_C]^{\mathrm{T}}$ |
| $\mathbf{v}_c$ | centre of $c$th cluster, prototype, $\mathbf{v}_c = [v_{c1}, \ldots, v_{cD}]$ |
| $\mathbf{Z}$ | weight matrix $[C \times D]$ |
| $z_{cd}$ | weight of $d$th attribute in $c$th cluster |
| $\hat{z}_{cd}$ | augmented weight of $d$th attribute in $c$th cluster, cf. Eq. (27) |
| $\beta_{ck}$ | typicality of the $k$th data item with respect to $c$th cluster |
| $f_k$ | global typicality of the $k$ data item, cf. Eq. (29) |
| $m$ | weighting exponent for memberships |
| $\phi$ | weighting exponent for weights |
| $h$ | loss function, cf. Eq. (31)–(37) |
| $e_{cdk}$ | residual of $d$th of $k$th datum from the centre of $c$th cluster |
| $\diamond$ | s-norm, cf. Eq. (29) |
| $\varkappa$ | data item's ordinal number, cf. Eq. (22) |

proposed [10,8,17,19,9,20]. Subspace FCM (fuzzy weighted C-means) is an algorithm that assigns weights (from interval [0, 1]) to attributes [18]. The algorithm is based on minimisation of a criterion function and calculates both fuzzy membership of data item to clusters and fuzzy weights for attributes in clusters. Automatic feature grouping fuzzy $k$-means (AFGFKM) algorithm groups attributes and assigns weights to the groups [9]. The membership of an data item to the group is crisp. The attribute weights in groups are soft.

One of the problems of the soft weight clustering algorithms is their vulnerability to noise and outliers [9]. The papers [12,13] present an ordering technique in clustering. The fuzzy C-ordered means algorithm (FCOM) [13] does not assign weights to attributes, but calculates typicality of each data item. The data items are ordered and their typicality is updated in each iteration of the clustering procedure. The distant items from all prototype centres have lower weights. Outliers and noise data items are assigned low typicality and do not distort the clustering process. Both algorithms are robust to outliers and even their high ratio does not distort the clustering results severely [12,13].

In the paper we propose a new fuzzy weighted C-ordered-means clustering algorithm. The algorithm finds clusters in fuzzy subspaces of the original task space. The algorithm assigns weights to dimensions (attributes) in each cluster. The weights are numbers from unit interval [0, 1]. To make this algorithm more robust to outliers and noise it incorporates the ordering technique.

In the paper we follow the general rule for symbols: the blackboard bold uppercase characters ($\mathbb{A}$) are used to denote the sets, uppercase italics ($A$) – the cardinality of sets, uppercase bolds ($\mathbf{A}$) matrices, lowercase bolds ($\mathbf{a}$) vectors, lowercase italics ($a$) scalars and set elements. Table 1 lists symbols used in the paper.

The paper is organised as follows: The novel fuzzy weighed c-ordered clustering algorithm is described in Sec. 2. The numerical experiments are presented in Sec. 3. Finally Sec. 4 summaries the paper.

## 2. Fuzzy weighted C-ordered-means clustering algorithm

The clustering algorithm minimises a criterion function $J$ defined as:

$$J(\mathbf{U}, \mathbf{V}, \mathbf{Z}) = \sum_{c=1}^{C} \sum_{i=1}^{X} \beta_{ci} (u_{ci})^m \sum_{d=1}^{D} (z_{cd})^{\phi} (x_{id} - v_{cd})^2 , \tag{1}$$

where $\mathbf{U}$ is a $C \times X$ membership matrix (for $C$ clusters and $X$ data items) whose each element $u_{cx} \in [0, 1]$ denotes the membership grade of the $x$th item to the $c$th cluster (group); $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_C]^{\mathrm{T}} \in \mathbb{R}^{C \times D}$ is a matrix of prototypes

$\mathbf{v}$ in which each row denotes centre (prototype) $\mathbf{v}_c = [v_{c1}, \dots, v_{cD}]$ of the $c$th cluster in a $D$ dimensional space; $\mathbf{Z} \in [0,1]^{C \times D}$ is a matrix of weights of attributes $d \in \{1, \dots, D\}$ in each cluster; $\beta_{ci} \in [0,1]$ denotes typicality of the $i$th data item with respect to the $c$th cluster; $m \in (0,1) \cup (1, \infty)$ is a weighting exponent for memberships (typical value is $m = 2$); $\phi \in (0,1) \cup (1, \infty)$ is a weighting exponent for weights.

The value of the criterion function is subject to minimisation with constraints:

$$\forall_{c \in \mathbb{C}} \sum_{d=1}^{D} z_{cd} = 1 \tag{2}$$

and

$$\forall_{k \in \mathbb{X}} \sum_{c=1}^{C} \beta_{ck} u_{ck} = f_k, \tag{3}$$

where $f_k$ is a global typicality of the $k$th data item (cf. formula (29)). To elaborate the formulae for membership and weight matrices we use Lagrange multipliers. The Lagrangian of (1) with constraints (2) and (3) is:

$$\begin{aligned} G(\mathbf{U}, \mathbf{Z}, \mathbf{V}) = \sum_{k=1}^{X} \Bigg( & \sum_{c=1}^{C} \beta_{ck} \left(u_{ck}\right)^m \sum_{d=1}^{D} \left(z_{cd}\right)^{\phi} \left(x_{kd} - v_{cd}\right)^2 + \\ & - \lambda_k \left[ \sum_{c=1}^{C} \beta_{ck} u_{ck} - f_k \right] - \lambda_c \left[ \sum_{d=1}^{D} z_{cd} - 1 \right] \Bigg), \end{aligned} \tag{4}$$

where $\lambda_k$ and $\lambda_c$ are Lagrange multipliers. Setting of the derivatives of the Lagrangian to zero results in:

$$\forall_{1 \leqslant k \leqslant X} \frac{\partial G(\mathbf{U}, \mathbf{Z}, \mathbf{V})}{\partial \lambda_k} = - \left[ \sum_{c=1}^{C} \beta_{ck} u_{ck} - f_k \right] = 0 \tag{5}$$

$$\forall_{1 \leqslant k \leqslant X} \frac{\partial G(\mathbf{U}, \mathbf{Z}, \mathbf{V})}{\partial \lambda_c} = - \left[ \sum_{d=1}^{D} z_{cd} - 1 \right] = 0 \tag{6}$$

$$\forall_{\substack{1 \leqslant k \leqslant X \\ 1 \leqslant s \leqslant C}} \frac{\partial G(\mathbf{U}, \mathbf{Z}, \mathbf{V})}{\partial u_{sk}} = \beta_{sk} m u_{sk}^{m-1} \sum_{d=1}^{D} \left(z_{sd}\right)^{\phi} \left(x_{kd} - v_{sd}\right)^2 - \lambda_k \beta_{sk} = 0 \tag{7}$$

$$\forall_{\substack{1 \leqslant s \leqslant C \\ 1 \leqslant a \leqslant D}} \frac{\partial G(\mathbf{U}, \mathbf{Z}, \mathbf{V})}{\partial z_{sa}} = \sum_{k=1}^{X} \beta_{sk} \left(u_{sk}\right)^m \phi \left(z_{sa}\right)^{\phi - 1} \left(x_{ka} - v_{sa}\right)^2 - \lambda_c = 0 \tag{8}$$

From (7) we get:

$$\beta_{sk} m u_{sk}^{m-1} \sum_{d=1}^{D} \left(z_{sd}\right)^{\phi} \left(x_{kd} - v_{sd}\right)^2 = \lambda_k \beta_{sk} \tag{9}$$

$$u_{sk}^{m-1} \sum_{d=1}^{D} \left(z_{sd}\right)^{\phi} \left(x_{kd} - v_{sd}\right)^2 = \frac{\lambda_k}{m} \tag{10}$$

$$u_{sk} = \left( \frac{\lambda_k}{m} \right)^{\frac{1}{m-1}} \left( \sum_{d=1}^{D} \left(z_{sd}\right)^{\phi} \left(x_{kd} - v_{sd}\right)^2 \right)^{\frac{1}{1-m}} \tag{11}$$

Combining (5) and (11) yields:

$$\sum_{c=1}^{C} \beta_{ck} \left( \frac{\lambda_k}{m} \right)^{\frac{1}{m-1}} \left( \sum_{d=1}^{D} \left(z_{cd}\right)^{\phi} \left(x_{kd} - v_{cd}\right)^2 \right)^{\frac{1}{1-m}} = f_k \tag{12}$$

$$\left(\frac{\lambda_k}{m}\right)^{\frac{1}{m-1}} \sum_{c=1}^{C} \beta_{ck} \left(\sum_{d=1}^{D} (z_{cd})^{\phi} (x_{kd} - v_{cd})^2\right)^{\frac{1}{1-m}} = f_k \tag{13}$$

Division of (11) by (13) results in formula:

$$\frac{u_{sk}}{f_k} = \frac{\left(\frac{\lambda_k}{m}\right)^{\frac{1}{m-1}} \left(\sum_{d=1}^{D} (z_{sd})^{\phi} (x_{kd} - v_{sd})^2\right)^{\frac{1}{1-m}}}{\left(\frac{\lambda_k}{m}\right)^{\frac{1}{m-1}} \sum_{c=1}^{C} \beta_{ck} \left(\sum_{d=1}^{D} (z_{cd})^{\phi} (x_{kd} - v_{cd})^2\right)^{\frac{1}{1-m}}} \tag{14}$$

$$u_{sk} = \frac{f_k \left(\sum_{d=1}^{D} (z_{sd})^{\phi} (x_{kd} - v_{sd})^2\right)^{\frac{1}{1-m}}}{\sum_{c=1}^{C} \beta_{ck} \left(\sum_{d=1}^{D} (z_{cd})^{\phi} (x_{kd} - v_{cd})^2\right)^{\frac{1}{1-m}}} \tag{15}$$

Similarly for attribute weights we start with (8):

$$\sum_{k=1}^{X} \beta_{sk} u_{sk}^{m} \phi z_{sa}^{\phi-1} (x_{ka} - v_{sa})^2 - \lambda_c = 0 \tag{16}$$

$$\phi z_{sa}^{\phi-1} \sum_{k=1}^{X} \beta_{sk} u_{sk}^{m} (x_{ka} - v_{sa})^2 = \lambda_c \tag{17}$$

For $\phi \neq 1$:

$$z_{sa} = \left(\frac{\lambda_c}{\phi}\right)^{\frac{1}{\phi-1}} \left[\sum_{k=1}^{X} \beta_{sk} u_{sk}^{m} (x_{ka} - v_{sa})^2\right]^{\frac{1}{1-\phi}} \tag{18}$$

Combining (6) and (18) yields:

$$\sum_{d=1}^{D} \left(\frac{\lambda_c}{\phi}\right)^{\frac{1}{\phi-1}} \left[\sum_{k=1}^{X} \beta_{sk} u_{sk}^{m} (x_{kd} - v_{sd})^2\right]^{\frac{1}{1-\phi}} = 1 \tag{19}$$

$$\left(\frac{\lambda_c}{\phi}\right)^{\frac{1}{\phi-1}} \sum_{d=1}^{D} \left[\sum_{k=1}^{X} \beta_{sk} u_{sk}^{m} (x_{kd} - v_{sd})^2\right]^{\frac{1}{1-\phi}} = 1 \tag{20}$$

Finally by division of (18) by (20) we get:

$$z_{sa} = \frac{\left[\sum_{k=1}^{X} \beta_{sk} u_{sk}^{m} (x_{ka} - v_{sa})^2\right]^{\frac{1}{1-\phi}}}{\sum_{d=1}^{D} \left[\sum_{k=1}^{X} \beta_{sk} u_{sk}^{m} (x_{kd} - v_{sd})^2\right]^{\frac{1}{1-\phi}}} \tag{21}$$

Before we present the pseudocode of the FWCOM algorithm, the ordering of data items should be discussed. For each data item in each cluster the typicality is calculated for each attribute separately. In order to do this data items are ordered by their distances from the cluster centre. The closest data item with respect to the $d$th attribute is labelled with an ordinal number 1, the most distant with $X$ (where $X$ is a number of data items). The value $\beta_{ckd}$ denotes typicality of the $k$th data item with respect to the $c$th cluster and $d$th attribute, and is calculated with function [13]

$$\beta_{ckd} = \max\left[\min\left(\frac{p_c X - \varkappa_{ckd}}{2 p_l X} + \frac{1}{2}; 1\right); 0\right] \tag{22}$$

called piecewise linearly weighted ordered weighted averaging (PLOWA) or

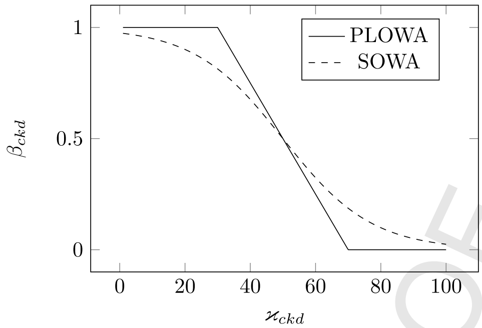



Fig. 1. The PLOWA and SOWA weighting functions for $X = 100$ and $p_a = 0.2$, $p_c = 0.5$, and $p_l = 0.2$.

$$\beta_{ckd} = \frac{1}{1 + \exp\left[\frac{2.944}{p_a X}\left(\varkappa_{ckd} - p_c X\right)\right]} \tag{23}$$

called sigmoidally weighted ordered weighted averaging (SOWA). In both functions $\varkappa_{ckd}$ stands for the index of the $k$th data item after reordering by the distance from the $c$th cluster with respect to the $d$th attribute. Both functions are presented in Fig. 1 for $X = 100$ and $p_a = 0.2$, $p_c = 0.5$, and $p_l = 0.2$.

Typicality of $i$th data item with respect to $c$th cluster is defined as a $t$-norm ($\star$) of typicalities of all attributes. Because the dimensions (attributes) are weighted we use the weighted T-norm [15]:

$$\begin{aligned}\beta_{ck} &= T\left(\beta_{ck1}, \beta_{ck2}, \ldots, \beta_{ckD}; z_{c1}^{\phi}, z_{c2}^{\phi}, \ldots, z_{cD}^{\phi}\right) = \\ &= T\left(1 - z_{c1}^{\phi}\left[1-\beta_{ck1}\right], 1 - z_{c2}^{\phi}\left[1-\beta_{ck2}\right], \ldots, 1 - z_{cD}^{\phi}\left[1-\beta_{ckD}\right]\right)\end{aligned} \tag{24}$$

In the algorithm we use product T-norm, thus:

$$\beta_{ck} = \prod_{d=1}^{D}\left(1 - z_{cd}^{\phi}\left[1-\beta_{ckd}\right]\right) \tag{25}$$

If all attributes have the highest importance (equal 1), their weights should not interfere with the typicality of a data item. In such a situation the weights are $z = \frac{1}{D}$. Unfortunately the nonunit value of weights interferes the typicality. If all dimensions have equal importance and typicality of each attribute is $\beta$, the formula (24) becomes:

$$\beta_{ck}(\beta) = \left[1 - \frac{1-\beta}{D^{\phi}}\right]^{D}. \tag{26}$$

If all attributes have minimal typicality (zero) the total typicality of the whole data item tends to one (with increase in the number of attributes), so there is no difference if the attributes are typical or not. Fig. 2 presents this phenomenon. It can be easily avoided by augmenting of the weights of the attributes [19]. The attribute weights for one data item are divided by the maximal values of them. This maximal value is always greater than zero. In the procedure all weights in the rule are scaled and the maximum weights become one:

$$\hat{z}_{cd} \leftarrow \frac{z_{cd}}{\max_{i\in[1,\ldots,D]} z_{ci}}, \tag{27}$$

where $\hat{z}_{cd}$ is an augmented weight of the $d$th attribute in the $c$th cluster. Thus instead of formula (24) we use the formula:

$$\beta_{ck} = T\left(1 - \hat{z}_{c1}^{\phi}\left[1-\beta_{ck1}\right], 1 - \hat{z}_{c2}^{\phi}\left[1-\beta_{ck2}\right], \ldots, 1 - \hat{z}_{cD}^{\phi}\left[1-\beta_{ckD}\right]\right) \tag{28}$$

The importance of the data item influences the localisation of the centres of clusters. The global typicality $f_k$ of $k$th data item is calculated with as a $s$-norm ($\diamond$) of typicalities of the data item in all clusters:

$$f_k = \beta_{1k} \diamond \beta_{2k} \diamond \ldots \diamond \beta_{Ck}. \tag{29}$$

We use the maximum operator as an $s$-norm operator.

Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001

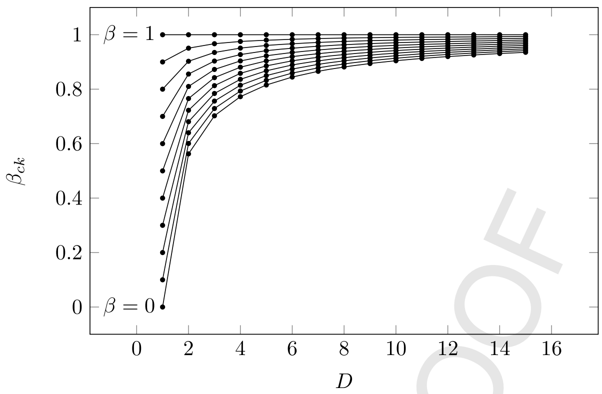



Fig. 2. Typicality for a data item (Eq. (25)) when all attributes have the equal typicality $\beta_{ck}$ in function of number of attributes ($D$) and attribute's weight exponent $\phi = 2$ without augmentation. If the weights of the attributes are not augmented the typicality of the data items tends to one independently whether the attributes have high or low typicality. The figure comprises eleven draws for values from $\beta = 0.0$ to 1.0 with 0.1 step. The lines are only to join the values for the same $\beta$'s. They have no physical meaning, because the number of attributes $D$ has only discrete values.

The centre of the $c$th cluster for the $d$th attribute is calculated with formula:

$$v_{cd} = \frac{\sum_{k=1}^{X} \beta_{ck} u_{ik}^{m} h\left(e_{ckd}\right) x_{kd}}{\sum_{k=1}^{X} \beta_{ck} u_{ik}^{m} h\left(e_{ckd}\right)} \tag{30}$$

where $e_{cdk} = |x_{kd} - v_{cd}|$ is a residual of the $d$th of the $k$th datum from the centre of the $c$th cluster, and $h : \mathbb{R} \to \mathbb{R}$ is a loss function defined in various ways [13]:

- essential loss function

$$h(x) = \begin{cases} 0, & x = 0 \\ 1, & x \neq 0 \end{cases} \tag{31}$$

- absolute (linear) loss function

$$h(x) = \begin{cases} 0, & x = 0 \\ |x|^{-1}, & x \neq 0 \end{cases} \tag{32}$$

- Huber (with parameter $\delta > 0$)

$$h(x) = \begin{cases} \delta^{-2}, & x = 0 \\ (\delta |x|)^{-1}, & x \neq 0 \end{cases} \tag{33}$$

- sigmoidal (with parameters $\alpha, \beta > 0$)

$$h(x) = \begin{cases} 0, & x = 0 \\ x^{-2} \left[1 + e^{(-\alpha(|x| - \beta)}\right]^{-1}, & x \neq 0 \end{cases} \tag{34}$$

- sigmoidal-linear (with parameters $\alpha, \beta > 0$)

$$h(x) = \begin{cases} 0, & x = 0 \\ |x|^{-1} \left[1 + e^{-\alpha(|x| - \beta)}\right]^{-1}, & x \neq 0 \end{cases} \tag{35}$$

- logarithmic

$$h(e) = \begin{cases} 0, & x = 0 \\ \frac{\log(1+x^2)}{x^2}, & x \neq 0 \end{cases} \tag{36}$$

- logarithmic-linear

$$h(e) = \begin{cases} 0, & x = 0 \\ \frac{\log(1+x^2)}{|x|}, & x \neq 0 \end{cases} \tag{37}$$

In experiments the following values of parameters are used [13]: $\alpha = 6.0$, $\beta = 1.0$, $\delta = 1.0$. Before the iterations the localisations of cluster centres are calculated with a formula

$$v_{cd} = \frac{\sum_{k=1}^{X} u_{ik}^{m} x_{kd}}{\sum_{k=1}^{X} u_{ik}^{m}}, \tag{38}$$

because the residuals $e$ in Eq. (30) are not known yet.

The pseudocode of the FWCOM (fuzzy weighted C-ordered mean) clustering algorithm is presented in Fig. 3.

## 3. Experiments

### 3.1. Datasets

#### 3.1.1. Artificial data set with varying number of outliers: 'art-outliers'

The data set 'art-outliers' is a six attribute data set with data item in two clusters with 100 items each:

**cluster 1** has its centre for 1st, 3rd, and 5th attributes at 3 (Gaussian distribution), other attributes are not valid (the values have uniform distribution);

**cluster 2** has its centre for 2nd, 4th, and 5th attributes at 7 (Gaussian distribution), other attributes are not valid (the values have uniform distribution).

The 5th attribute is important for both clusters, the 6th attribute is not important in either cluster. The outliers at $(0, 0, 0, 0, 0, 0)$ are added to the clusters described above. The number of outliers is 0 to 100 with step 5. The data items were generated 50 times for each number of outliers.

#### 3.1.2. Artificial data set with background noise: 'art-noise'

In this data set the clusters are defined in the same way as in the data set 'art-outliers'. The only difference is that the data set 'art-noise' has not got the outliers as the 'art-outliers', but instead the uniform noise from interval $[0, 10]$ is added to the data.

#### 3.1.3. Methane concentration

The data set contains the real life measurements of air parameters in a coal mine in Upper Silesia (Poland). The parameters (measured in 10 second intervals) are: AN31 – the flow of air in the shaft, AN32 – the flow of air in the adjacent shaft, MM32 – concentration of methane ($CH_4$), production of coal. To the tuples the 10-minute sums of measurements of AN31, AN32, MM32 are added as dynamic attributes [16]. The task is to predict the concentration of the methane in 10 minutes. The data is divided into train set (499 tuples) and test set (523 tuples).

#### 3.1.4. Death rate

The data represent the tuples containing information on various factors, the task is to estimate the death rate [21]. The first attribute (item index) is excluded from the dataset. The following attributes are: average annual precipitation, average January temperature, average July temperature, size of the population older than 65, number of members per household, years of schooling for persons over 22, households with fully equipped kitchens, population per square mile, size of the nonwhite population, number of office workers, families with an income $< \$3000$, hydrocarbon pollution index, nitric oxide pollution index, sulphur dioxide pollution index, degree of atmospheric moisture, death rate. The data set can be downloaded from public repository.[1]

[1] http://orion.math.iastate.edu/burkardt/data/regression/x28.txt.

Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001

```
procedure FWCOM (X, c, m, ϕ, N)
// X : matrix of data items
// c : number of clusters
// m : weighting exponent for memberships
// ϕ : weighting exponent for dimensions
// N : number of iterations

   X := number of data item in X;

   initialise all β's with 1's;
   calculate typicality of data items with (29);
   initialise U with random numbers and normalise with constraint (3);
   initialise Z with random numbers and normalise with constrain (2);
   calculate cluster prototypes V with (38);

   for iter := 1 to N do
      update U with (15) and normalise with (3);
      update attribute weights with (21) with constraint (2);

      // update centres of prototypes (clusters):
      for c := 1 to C do // for each cluster
         for d := 1 to D do // for each attribute
            for k := 1 to X do // for each data item
               e_cdk := |x_kd − v_cd|;   // calculate residuals
            end for;
            sort the residuals;
            label each residual with an ordinal number ϰ in sorted sequence;
            for k := 1 to X do // for each residual
               calculate typicality β_ckd with (22),(23), or uniform weighting
            end for;
         end for; // for each attribute

         for k := 1 to X do // for each data item
            calculate typicality β_ck with (28);
         end for;
         update prototype with (30);
      end for; // for each cluster

      for k := 1 to X do // for each data item
         update typicality of data item with (29);
      end for;
   end for;

   return U, V, Z;
end procedure;
```

Fig. 3. Pseudocode of the Fuzzy Weighted C-Ordered Means (FWCOM) clustering algorithm.



### 3.1.5. *Concrete compressive strength*

The 'Concrete' set is a real life data set describing the parameters of the concrete sample and its strength [23]. The attributes are: cement ratio, amount of blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, fine aggregate, age; the decision attribute is the concrete compressive strength. The original data set can be downloaded[2] from public repository [6].

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

Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001

### 3.2. Numerical experiments

In the experiment we used clustering algorithms: FCM [5], fuzzy weighted c-means, FWCM (subspace FCM) [17], AGFKM [9], FCOM [13], and the proposed FWCOM algorithm.

The FWCOM algorithm has a weight parameter $\phi$, that needs calibration. For this the artificial data sets 'art-outliers' and 'art-noise' were used. The cluster centres in the both data sets can be gathered in a form of a matrix $\mathbf{V}_d$ (matrix of desired values – subscript $_d$)

$$\mathbf{V}_d = \begin{bmatrix} 3 & 0 & 3 & 0 & 3 & 0 \\ 0 & 7 & 0 & 7 & 7 & 0 \end{bmatrix}, \tag{39}$$

where each row represents one cluster and each column one attribute. The desired values of weight in form of a matrix $\mathbf{Z}_d$:

$$\mathbf{Z}_d = \begin{bmatrix} \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix} \tag{40}$$

with the same meaning of rows and columns as $\mathbf{V}_d$. Each row of matrix $\mathbf{Z}_d$ fulfils the constraint (2).

In the same way matrices $\mathbf{V}_e$ and $\mathbf{Z}_e$ represent the cluster centres and attributes weights elaborated (subscript $_e$) by the FWCM or FWCOM algorithms. The index of clustering error $E_v$ for cluster centres is calculated as a Frobenius norm $\|\cdot\|_F$ of difference of the Hadamard products of matrices of cluster centres and weights:

$$E_v = \|\mathbf{V}_d \circ \mathbf{Z}_d - \mathbf{V}_e \circ \mathbf{Z}_e\|_F\,, \tag{41}$$

where $\circ$ denotes a Hadamard product of matrices defined as:

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \circ \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \dots & b_{mn} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & \dots & a_{1n}b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1}b_{m1} & \dots & a_{mn}b_{mn} \end{bmatrix} \tag{42}$$

and Frobenius norm of a matrix $\mathbf{A} = [a]^{m\times n}$ with $m$ rows and $n$ columns is defined as

$$\|\mathbf{A}\|_F = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2. \tag{43}$$

For values of weights of attributes we define the error index $E_z$:

$$E_z = \|\mathbf{Z}_d - \mathbf{Z}_e\|_F \tag{44}$$

Lower values of measures $E_v$ and $E_z$ denote more precise match of centre localisation and weights.

The algorithm assigns typicalities to data items. We use a name *informative data items* for regular data items, and *noninformative data items* for outliers and noise in the data set. The outliers should have low typicalities, whereas informative data should have high typicalities. The accuracy of typicalities can be measured separately for informative $\xi_i$ and noninformative data items (noise or outliers) $\xi_o$ with a complement to one of root mean square error of typicalities defined as

$$\xi_i = 1 - \sqrt{\frac{1}{K_i}\sum_{k}^{K_i}(f_k - f_i)^2}, \tag{45}$$

$$\xi_n = 1 - \sqrt{\frac{1}{K_n}\sum_{k}^{K_n}(f_k - f_n)^2}, \tag{46}$$

where $K_i$ ($K_n$) is a number of informative (noninformative) items, $f_k$ stands for typicality of the $k$th item, and $f_i$, $f_n$ are typicalities of the data classes: $f_i = 1$ for informative data items, $f_n = 0$ for outliers or noise. If typicalities of data items are elaborated perfectly (0 for outliers and 1 for data items) the value of accuracy is $\xi_i = \xi_n = 1$. When the typicalities are totally opposite, then $\xi_i = \xi_n = 0$.

Because the algorithm uses random values in initialisation of membership and weight matrices, the experiments were repeated 50 times.

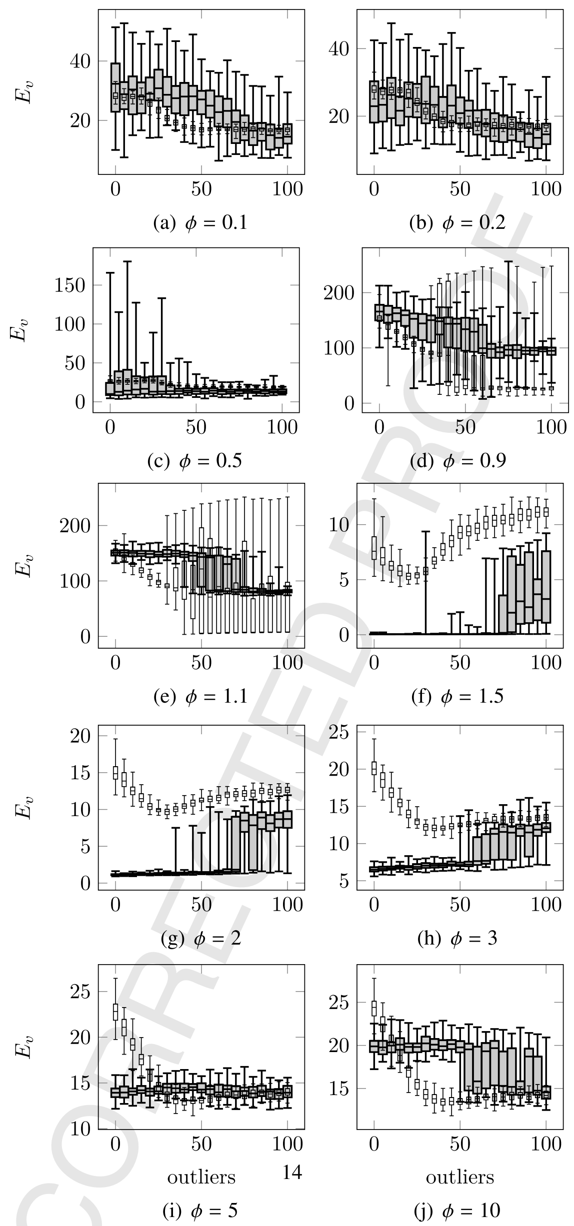



Fig. 4. The box plot of error index for centre localisation ($E_v$) for the ‘art-outliers’ dataset elaborated by FWCM (narrow boxes) and FWCOM (grey wide boxes) with essential loss function.



### 3.2.1. *Artificial dataset with varying number of outliers: ‘art-outliers’*

The experiments aim at testing the influence of parameter $\phi$ and loss function on the localisation of elaborated clusters and on the values of weights of attributes in clusters. The number of clusters is an input parameter of the FWCOM algorithm. For the ‘art-outliers’ data set number of clusters $C = 2$.

The influence of weighting parameter $\phi$ on the results of the clustering has been tested for the weight parameter $\phi \in \{0.1, 0.2, 0.5, 0.9, 1.1, 1.2, 1.5, 2.0, 3.0, 5.0, 10.0\}$. The influence of the weight parameter $\phi$ has been tested for all loss functions described in Sec. 2. To save the space of the article we provide the complete box plots only for the essential loss function. The results of experiments for other tested loss functions lead to similar conclusion. In experiments the SOWA (Eq. (23)) operator is used.

Figs. 4(a)–4(j) present the box plots of error index $E_v$ for FWCM (narrow boxes) and FWCOM (wide grey boxes) for various values of the $\phi$ parameter for essential loss function. The rectangular of the box plot represents the in-

Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001

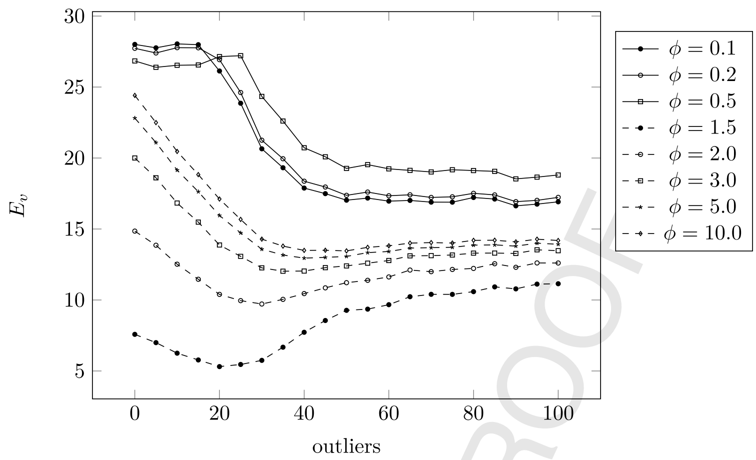



Fig. 5. The median of error index for centre localisation ($E_v$) for the 'art-outliers' dataset elaborated by FWCM algorithm.

Table 2
The median of error index for centre localisation ($E_v$) for the 'art-outliers' dataset elaborated by FWCM algorithm.

| Outliers | Weight parameter $\phi$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0.1 | 0.2 | 0.5 | 1.5 | 2 | 3 | 5 | 10 |
| 0 | 28.004 | 27.728 | 26.839 | 7.579 | 14.851 | 20.002 | 22.815 | 24.410 |
| 5 | 27.762 | 27.400 | 26.390 | 6.995 | 13.856 | 18.606 | 21.087 | 22.495 |
| 10 | 28.038 | 27.773 | 26.533 | 6.252 | 12.520 | 16.827 | 19.147 | 20.462 |
| 15 | 27.984 | 27.764 | 26.556 | 5.775 | 11.458 | 15.483 | 17.627 | 18.830 |
| 20 | 26.130 | 26.929 | 27.144 | 5.307 | 10.393 | 13.873 | 15.952 | 17.111 |
| 25 | 23.859 | 24.614 | 27.209 | 5.457 | 9.953 | 13.072 | 14.724 | 15.675 |
| 30 | 20.649 | 21.254 | 24.339 | 5.746 | 9.716 | 12.263 | 13.575 | 14.286 |
| 35 | 19.315 | 19.949 | 22.605 | 6.673 | 10.044 | 12.027 | 13.166 | 13.794 |
| 40 | 17.883 | 18.366 | 20.723 | 7.724 | 10.449 | 12.037 | 12.950 | 13.492 |
| 45 | 17.491 | 17.956 | 20.087 | 8.553 | 10.855 | 12.277 | 13.001 | 13.501 |
| 50 | 17.029 | 17.372 | 19.267 | 9.270 | 11.218 | 12.399 | 13.068 | 13.458 |
| 55 | 17.178 | 17.603 | 19.535 | 9.354 | 11.385 | 12.592 | 13.335 | 13.708 |
| 60 | 16.963 | 17.341 | 19.237 | 9.668 | 11.623 | 12.780 | 13.416 | 13.815 |
| 65 | 17.012 | 17.409 | 19.129 | 10.229 | 12.108 | 13.109 | 13.655 | 14.006 |
| 70 | 16.892 | 17.224 | 19.018 | 10.398 | 11.995 | 13.123 | 13.685 | 14.042 |
| 75 | 16.889 | 17.267 | 19.174 | 10.394 | 12.150 | 13.167 | 13.711 | 14.011 |
| 80 | 17.219 | 17.513 | 19.116 | 10.587 | 12.222 | 13.299 | 13.862 | 14.199 |
| 85 | 17.110 | 17.401 | 19.061 | 10.921 | 12.553 | 13.303 | 13.881 | 14.209 |
| 90 | 16.631 | 16.925 | 18.530 | 10.782 | 12.302 | 13.278 | 13.794 | 14.091 |
| 95 | 16.751 | 17.018 | 18.653 | 11.118 | 12.609 | 13.534 | 14.000 | 14.282 |
| 100 | 16.914 | 17.227 | 18.805 | 11.144 | 12.604 | 13.472 | 13.924 | 14.189 |

terquartile range. The horizontal bar in the rectangular stands for the median. Upper and lower whiskers denote maximum and minimum respectively. For easier comparison the median values of the error index $E_v$ for the FWCM algorithm are presented in Fig. 5 and Table 2, for the FWCOM algorithm—in Fig. 6 and Table 3. For $\phi < 1$ the results are significantly poorer than for higher values. The lowest values of $E_v$ are elaborated with $\phi = 1.5$. The second lowest group of results are for $\phi = 2$. Although for $\phi = 1.5$ the median of the error index $E_v$ for cluster centres has the lowest values, the results have higher dispersion than $\phi = 2$. The boxes in Fig. 4(f) for $\phi = 1.5$ show wider interquartile ranges than in Fig. 4(g) for $\phi = 2$. These conclusions are also valid for other loss function (although the results are not presented in the paper). The value $\phi = 1$ is illegal. Fig. 4(d) and 4(e) show the results elaborated for neighbour values $\phi = 0.9$ and $\phi = 1.1$ respectively. It can be easily seen that the results are severely distorted for both FWCM and FWCOM clustering algorithms.

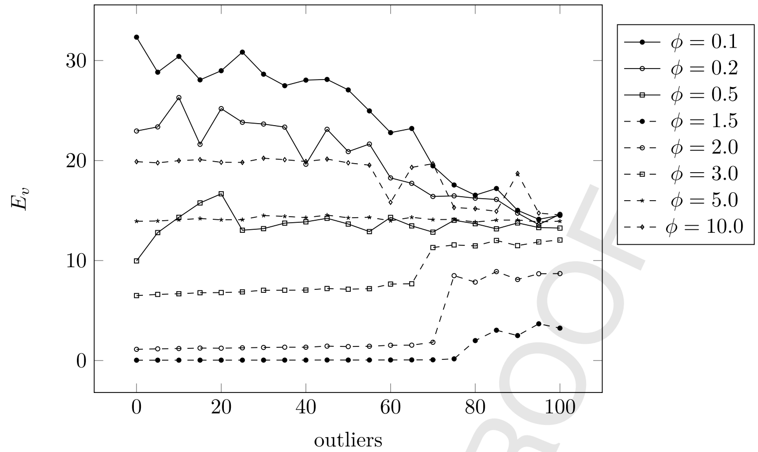



Fig. 6. The median of error index for centre localisation ($E_v$) for the 'art-outliers' dataset elaborated by FWCOM algorithm, loss function: essential.

Table 3
The median of error index for centre localisation ($E_v$) for the 'art-outliers' dataset elaborated by FWCOM algorithm, loss function: essential.

| Outliers | Weight parameter $\phi$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0.1 | 0.2 | 0.5 | 1.5 | 2 | 3 | 5 | 10 |
| 0 | 32.332 | 22.951 | 9.962 | 0.035 | 1.124 | 6.496 | 13.929 | 19.885 |
| 5 | 28.828 | 23.361 | 12.796 | 0.038 | 1.171 | 6.616 | 13.955 | 19.769 |
| 10 | 30.409 | 26.295 | 14.323 | 0.039 | 1.203 | 6.676 | 14.076 | 19.985 |
| 15 | 28.062 | 21.618 | 15.758 | 0.041 | 1.249 | 6.781 | 14.201 | 20.089 |
| 20 | 28.978 | 25.190 | 16.674 | 0.041 | 1.236 | 6.798 | 14.072 | 19.822 |
| 25 | 30.835 | 23.822 | 13.033 | 0.043 | 1.274 | 6.864 | 14.085 | 19.812 |
| 30 | 28.629 | 23.647 | 13.189 | 0.044 | 1.309 | 7.030 | 14.499 | 20.231 |
| 35 | 27.480 | 23.341 | 13.743 | 0.045 | 1.336 | 7.035 | 14.417 | 20.083 |
| 40 | 28.038 | 19.626 | 13.868 | 0.047 | 1.334 | 7.047 | 14.290 | 19.905 |
| 45 | 28.108 | 23.122 | 14.223 | 0.055 | 1.441 | 7.203 | 14.541 | 20.144 |
| 50 | 27.060 | 20.894 | 13.650 | 0.051 | 1.398 | 7.134 | 14.271 | 19.767 |
| 55 | 24.961 | 21.646 | 12.895 | 0.054 | 1.431 | 7.188 | 14.324 | 19.535 |
| 60 | 22.789 | 18.262 | 14.313 | 0.059 | 1.531 | 7.647 | 13.977 | 15.823 |
| 65 | 23.203 | 17.719 | 13.464 | 0.066 | 1.541 | 7.677 | 14.321 | 19.315 |
| 70 | 19.467 | 16.407 | 12.839 | 0.075 | 1.829 | 11.30 | 14.091 | 19.678 |
| 75 | 17.553 | 16.467 | 14.026 | 0.170 | 8.489 | 11.56 | 14.112 | 15.301 |
| 80 | 16.537 | 16.234 | 13.681 | 1.997 | 7.846 | 11.46 | 13.854 | 15.184 |
| 85 | 17.203 | 16.107 | 13.170 | 3.034 | 8.895 | 12.00 | 14.042 | 14.922 |
| 90 | 15.022 | 14.740 | 13.763 | 2.499 | 8.107 | 11.49 | 14.051 | 18.705 |
| 95 | 14.125 | 13.592 | 13.297 | 3.676 | 8.679 | 11.86 | 13.866 | 14.745 |
| 100 | 14.512 | 14.624 | 13.247 | 3.242 | 8.697 | 12.05 | 13.932 | 14.558 |

Fig. 7 and Table 4 present the medians of error measure for attribute weights $E_z$ for various values of the parameter $\phi$ for essential loss function. The lowest values of $E_z$ are elaborated for $\phi = 2$. The interesting fact is a rapid increase of the error measure $E_z$ for $\phi = 1.5$, for which the error measure for centre localisation $E_v$ has the lowest values.

The influence of the loss function is presented in following figures and tables. Fig. 8 and Table 5 present the medians of the error index $E_v$ for various loss functions. Fig. 9 and Table 6 present the error index $E_z$ for various loss functions. The interesting fact is the moderate increase of error measure of centre localisation with the increase of number of outliers. This phenomenon can be observed for the linear, sigmoidal, and sigmoidal linear loss functions (cf. Table 5). It is worth noticing that the number of informative data items is 200 and for 95 outliers (almost one third of all data) the error measure is 1.663, whereas for the data set with no outliers it is 1.164. The other loss functions have significant increase in $E_v$ and $E_z$ for higher ratios of outliers.

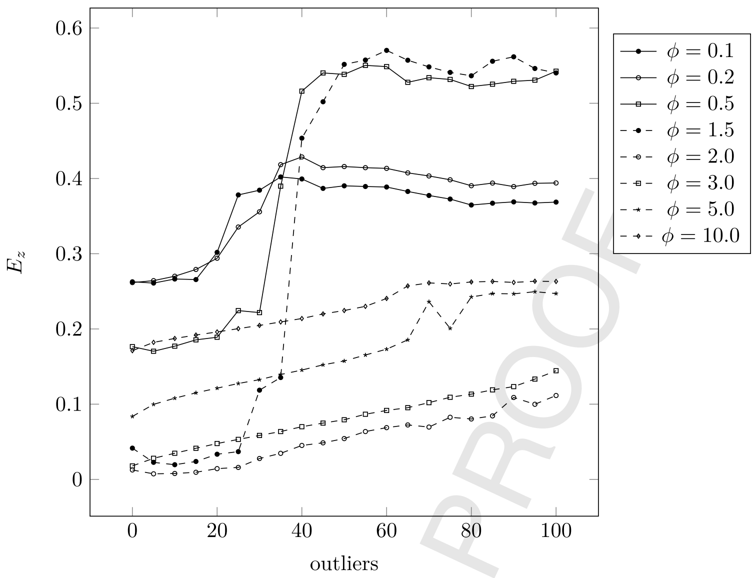



Fig. 7. Medians of error index for weights ($E_z$) for the ‘art-outliers’ dataset elaborated by the FWCOM algorithm with essential loss function.

Table 4
Medians of error index for weights ($E_z$) for the ‘art-outliers’ dataset elaborated by the FWCOM algorithm with essential loss function.

| Outliers | Weight parameter $\phi$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0.1 | 0.2 | 0.5 | 1.5 | 2 | 3 | 5 | 10 |
| 0 | 0.2626 | 0.2616 | 0.1764 | 0.0415 | 0.0125 | 0.0179 | 0.0836 | 0.1711 |
| 5 | 0.2609 | 0.2642 | 0.1703 | 0.0226 | 0.0074 | 0.0280 | 0.0997 | 0.1821 |
| 10 | 0.2663 | 0.2700 | 0.1771 | 0.0195 | 0.0078 | 0.0347 | 0.1079 | 0.1873 |
| 15 | 0.2656 | 0.2790 | 0.1855 | 0.0238 | 0.0094 | 0.0413 | 0.1151 | 0.1919 |
| 20 | 0.3018 | 0.2940 | 0.1889 | 0.0334 | 0.0143 | 0.0476 | 0.1212 | 0.1960 |
| 25 | 0.3781 | 0.3355 | 0.2243 | 0.0369 | 0.0159 | 0.0531 | 0.1275 | 0.2004 |
| 30 | 0.3844 | 0.3557 | 0.2217 | 0.1185 | 0.0277 | 0.0584 | 0.1325 | 0.2047 |
| 35 | 0.4020 | 0.4184 | 0.3897 | 0.1354 | 0.0346 | 0.0636 | 0.1394 | 0.2092 |
| 40 | 0.3993 | 0.4286 | 0.5160 | 0.4535 | 0.0451 | 0.0700 | 0.1453 | 0.2138 |
| 45 | 0.3868 | 0.4143 | 0.5402 | 0.5019 | 0.0486 | 0.0748 | 0.1523 | 0.2199 |
| 50 | 0.3902 | 0.4159 | 0.5384 | 0.5517 | 0.0542 | 0.0792 | 0.1574 | 0.2245 |
| 55 | 0.3893 | 0.4143 | 0.5503 | 0.5574 | 0.0636 | 0.0866 | 0.1653 | 0.2300 |
| 60 | 0.3886 | 0.4134 | 0.5487 | 0.5702 | 0.0688 | 0.0916 | 0.1731 | 0.2405 |
| 65 | 0.3827 | 0.4074 | 0.5278 | 0.5571 | 0.0723 | 0.0953 | 0.1855 | 0.2571 |
| 70 | 0.3774 | 0.4033 | 0.5339 | 0.5483 | 0.0695 | 0.1020 | 0.2361 | 0.2611 |
| 75 | 0.3726 | 0.3983 | 0.5316 | 0.5410 | 0.0825 | 0.1091 | 0.2008 | 0.2598 |
| 80 | 0.3647 | 0.3903 | 0.5222 | 0.5365 | 0.0803 | 0.1133 | 0.2425 | 0.2624 |
| 85 | 0.3671 | 0.3939 | 0.5253 | 0.5558 | 0.0844 | 0.1190 | 0.2470 | 0.2631 |
| 90 | 0.3688 | 0.3892 | 0.5291 | 0.5617 | 0.1089 | 0.1233 | 0.2466 | 0.2618 |
| 95 | 0.3674 | 0.3935 | 0.5307 | 0.5462 | 0.0998 | 0.1333 | 0.2494 | 0.2634 |
| 100 | 0.3686 | 0.3940 | 0.5425 | 0.5403 | 0.1115 | 0.1444 | 0.2470 | 0.2631 |

The Figs. 10(a) and 10(b) present the box plots of values of weights for the first and second cluster for data set with 15 outliers. These figures are typical for all tested values of outliers, so we do not present plots for all tested values. The grey narrow boxes present the values elaborated by the FWCM algorithms, the black wide ones – by the FWCOM algorithms. The values elaborated by the latter algorithm show that this algorithm can elaborate the weights of the attributes more precisely.

The FWCOM algorithm elaborates the typicalities ($f$, Eq. (29)) of data items. Fig. 11 presents the histogram of typicalities for the ‘art-outliers’ data set with 100 outliers. All informative data items are assigned with high values

Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001

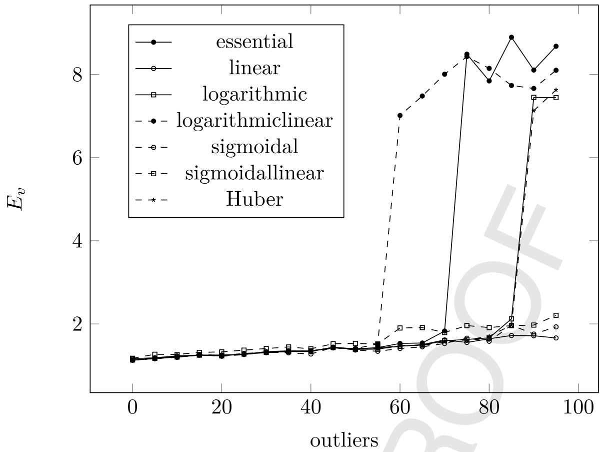



Fig. 8. Medians of error index for centre localisation ($E_v$) for the ‘art-outliers’ dataset for weight parameter $\phi = 2$.

Table 5
The median of error index for centre localisation ($E_v$) for the ‘art-outliers’ dataset elaborated by FWCOM algorithm, $\phi = 2$.

| Outliers | Loss function | | | | | | |
|---|---|---|---|---|---|---|---|
| | lin | ess | log | log-lin | sig | sig-lin | Huber |
| 0 | 1.163 | 1.124 | 1.133 | 1.130 | 1.136 | 1.174 | 1.129 |
| 5 | 1.186 | 1.171 | 1.171 | 1.171 | 1.163 | 1.269 | 1.171 |
| 10 | 1.231 | 1.203 | 1.205 | 1.195 | 1.204 | 1.267 | 1.218 |
| 15 | 1.251 | 1.249 | 1.249 | 1.263 | 1.261 | 1.314 | 1.254 |
| 20 | 1.253 | 1.236 | 1.232 | 1.240 | 1.221 | 1.328 | 1.238 |
| 25 | 1.271 | 1.274 | 1.277 | 1.290 | 1.264 | 1.369 | 1.271 |
| 30 | 1.334 | 1.309 | 1.315 | 1.317 | 1.320 | 1.408 | 1.325 |
| 35 | 1.353 | 1.336 | 1.334 | 1.337 | 1.305 | 1.446 | 1.339 |
| 40 | 1.349 | 1.334 | 1.352 | 1.349 | 1.276 | 1.400 | 1.344 |
| 45 | 1.423 | 1.441 | 1.433 | 1.431 | 1.436 | 1.524 | 1.425 |
| 50 | 1.409 | 1.398 | 1.388 | 1.406 | 1.372 | 1.530 | 1.389 |
| 55 | 1.418 | 1.431 | 1.394 | 1.519 | 1.342 | 1.518 | 1.412 |
| 60 | 1.461 | 1.531 | 1.471 | 7.015 | 1.409 | 1.902 | 1.485 |
| 65 | 1.503 | 1.541 | 1.489 | 7.483 | 1.449 | 1.912 | 1.481 |
| 70 | 1.615 | 1.829 | 1.595 | 8.007 | 1.530 | 1.795 | 1.568 |
| 75 | 1.554 | 8.489 | 1.626 | 8.431 | 1.656 | 1.960 | 1.616 |
| 80 | 1.643 | 7.846 | 1.664 | 8.147 | 1.584 | 1.912 | 1.704 |
| 85 | 1.722 | 8.895 | 2.121 | 7.736 | 1.971 | 1.958 | 1.969 |
| 90 | 1.716 | 8.107 | 7.449 | 7.664 | 1.756 | 1.972 | 7.130 |
| 95 | 1.662 | 8.679 | 7.444 | 8.102 | 1.929 | 2.205 | 7.629 |

of typicality, whereas all outliers have very low typicalities. This proves that the algorithm can correctly distinguish informative items from outliers.

Fig. 11 presents the histogram of typicality of data item in ‘art-outliers’ data set. The values of typicality are elaborated with FWCOM algorithm with essential loss function. The data item on the left side of the histogram represent outliers, the item on the right side – the informative data items. The algorithm assigns typicalities from interval [0, 1] to data items, but does not determine a value of typicality that separates outliers from informative data items. In the ‘art-outliers’ data set the maximum typicality value of an outlier is 0.129021 and minimal typicality of an informative data item is 0.637216. In this case every value for the interval $(0.129021, 0.637216)$ separates informative data from outliers. This can be explained by the fact that outliers lie far from cluster centres thus they have low typicalities. For data presented in Fig. 11 the accuracy (defined with formulae (45) and (46)) for informative data items is $\xi_i = 0.9052$, for outliers: $\xi_n = 0.9637$.

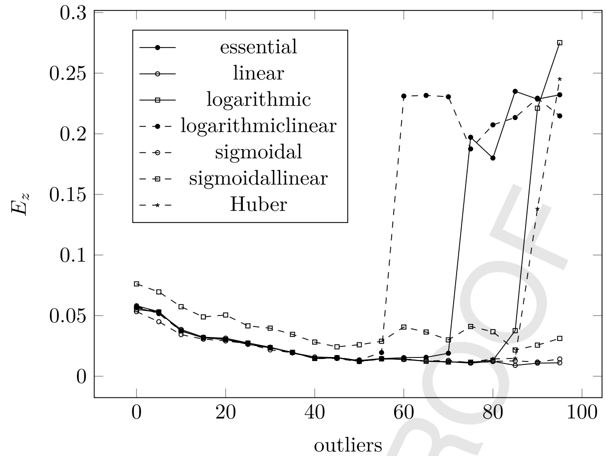



Fig. 9. Medians of error index for weights ($E_z$) for the ‘art-outliers’ dataset for weight parameter $\phi = 2$ elaborated with the FWCOM algorithm.

Table 6
Medians of error index for weights ($E_z$) for the ‘art-outliers’ dataset for weight parameter $\phi = 2$ elaborated with FWCOM algorithm.

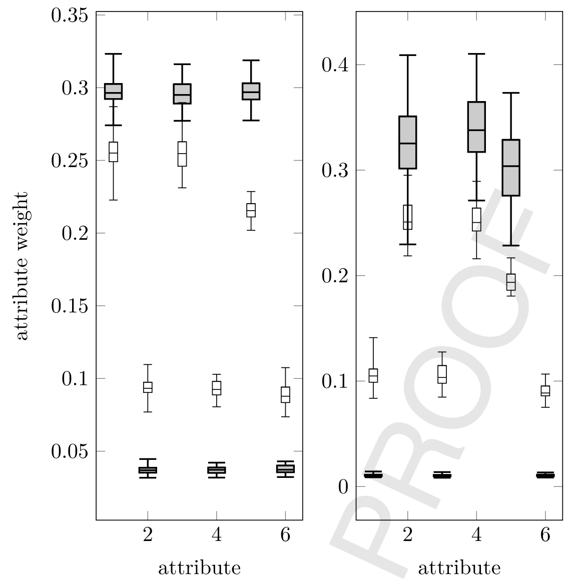
| Outliers | Loss function | | | | | | |
|---|---|---|---|---|---|---|---|
| | lin | ess | log | log-lin | sig | sig-lin | Huber |
| 0 | 0.056 | 0.058 | 0.055 | 0.057 | 0.053 | 0.076 | 0.058 |
| 5 | 0.052 | 0.053 | 0.053 | 0.051 | 0.044 | 0.069 | 0.052 |
| 10 | 0.038 | 0.037 | 0.037 | 0.037 | 0.034 | 0.057 | 0.037 |
| 15 | 0.032 | 0.031 | 0.031 | 0.031 | 0.030 | 0.048 | 0.031 |
| 20 | 0.031 | 0.030 | 0.030 | 0.030 | 0.029 | 0.050 | 0.030 |
| 25 | 0.027 | 0.026 | 0.027 | 0.026 | 0.026 | 0.041 | 0.026 |
| 30 | 0.023 | 0.023 | 0.023 | 0.023 | 0.021 | 0.039 | 0.023 |
| 35 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.034 | 0.019 |
| 40 | 0.015 | 0.014 | 0.014 | 0.014 | 0.014 | 0.028 | 0.014 |
| 45 | 0.015 | 0.015 | 0.015 | 0.015 | 0.014 | 0.024 | 0.015 |
| 50 | 0.013 | 0.012 | 0.012 | 0.013 | 0.012 | 0.025 | 0.012 |
| 55 | 0.014 | 0.014 | 0.014 | 0.019 | 0.014 | 0.028 | 0.014 |
| 60 | 0.013 | 0.015 | 0.014 | 0.231 | 0.014 | 0.040 | 0.014 |
| 65 | 0.012 | 0.015 | 0.012 | 0.231 | 0.012 | 0.036 | 0.012 |
| 70 | 0.012 | 0.018 | 0.011 | 0.230 | 0.013 | 0.030 | 0.011 |
| 75 | 0.010 | 0.197 | 0.011 | 0.187 | 0.011 | 0.041 | 0.011 |
| 80 | 0.012 | 0.179 | 0.013 | 0.207 | 0.012 | 0.036 | 0.014 |
| 85 | 0.008 | 0.235 | 0.037 | 0.213 | 0.012 | 0.021 | 0.014 |
| 90 | 0.010 | 0.228 | 0.221 | 0.229 | 0.011 | 0.025 | 0.138 |
| 95 | 0.011 | 0.232 | 0.275 | 0.214 | 0.014 | 0.031 | 0.245 |



Table 7
The median of error index for centre localisation ($E_v$) for the ‘art-noise’ dataset elaborated by FWCOM algorithm, loss function: essential.

| Outliers | Weight parameter $\phi$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0.1 | 0.2 | 0.5 | 1.5 | 2 | 3 | 5 | 10 |
| 0 | 29.961 | 20.502 | 12.652 | 0.037 | 1.152 | 6.5895 | 14.043 | 19.984 |
| 25 | 28.056 | 22.309 | 12.583 | 0.084 | 2.287 | 9.2242 | 16.117 | 20.908 |
| 50 | 31.548 | 26.096 | 22.348 | 0.143 | 3.221 | 10.834 | 17.196 | 21.536 |
| 75 | 29.393 | 27.314 | 16.833 | 0.200 | 3.957 | 12.272 | 18.542 | 22.354 |
| 100 | 29.795 | 24.205 | 18.537 | 0.266 | 4.643 | 12.890 | 18.778 | 22.359 |
| 125 | 29.324 | 26.946 | 25.103 | 0.382 | 5.256 | 13.867 | 19.254 | 22.659 |
| 150 | 31.091 | 28.416 | 21.100 | 0.521 | 6.125 | 14.773 | 20.015 | 23.073 |
| 175 | 31.277 | 28.912 | 33.704 | 0.580 | 6.592 | 15.465 | 20.609 | 23.335 |
| 200 | 30.961 | 30.046 | 20.975 | 0.731 | 7.361 | 16.074 | 20.696 | 23.277 |

Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001




(a) 1st cluster (b) 2nd cluster

Fig. 10. The box plots for values of weights in the clusters of the ‘art-outliers’ dataset with 15 outliers elaborated by FWCM (narrow boxes) and FWCOM (grey wide boxes) algorithms for $\phi = 2$, loss function of FWCOM: essential.

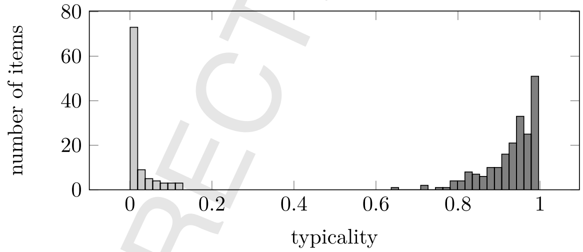



Fig. 11. Histogram of typicality of data item from the ‘art-outliers’ data set with 100 outliers. The typicalities are elaborated with the FWCOM algorithm with essential loss function.

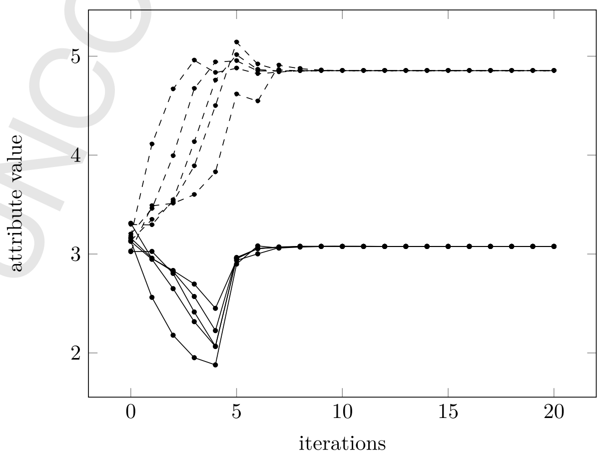



Fig. 12. Changes of elaborated values of cluster centre for the first attribute of the ‘art-outliers’ data set. Results of five clustering are shown. The first and second clusters are represented by solid and dashed lines. The membership matrix **U** is initialised with random numbers.



Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001

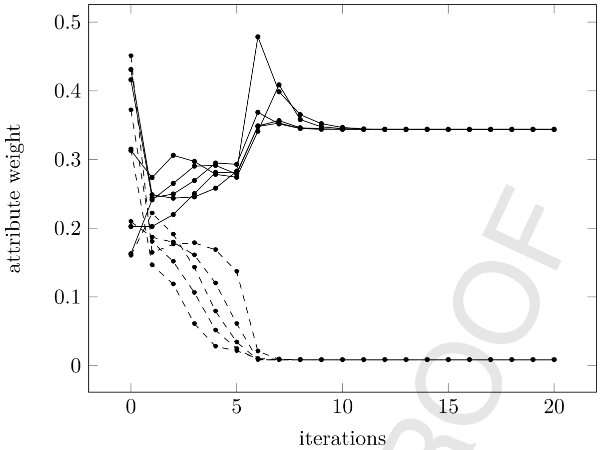



Fig. 13. Changes of elaborated values of attribute weights for the first attribute of the 'art-outliers' data set. Results of five clustering are shown. The first and second clusters are represented by solid and dashed lines. The attribute weight matrix **Z** is initialised with random numbers.

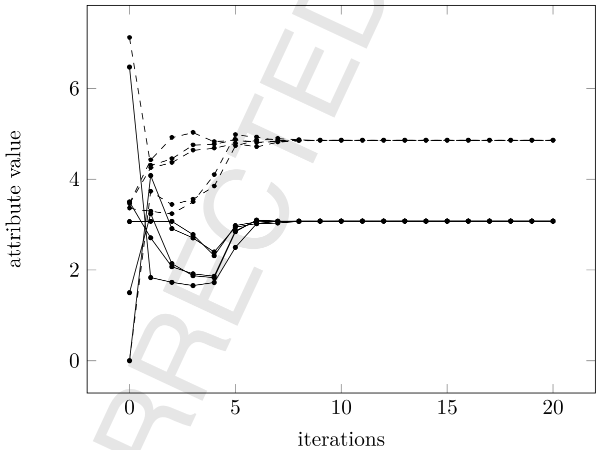
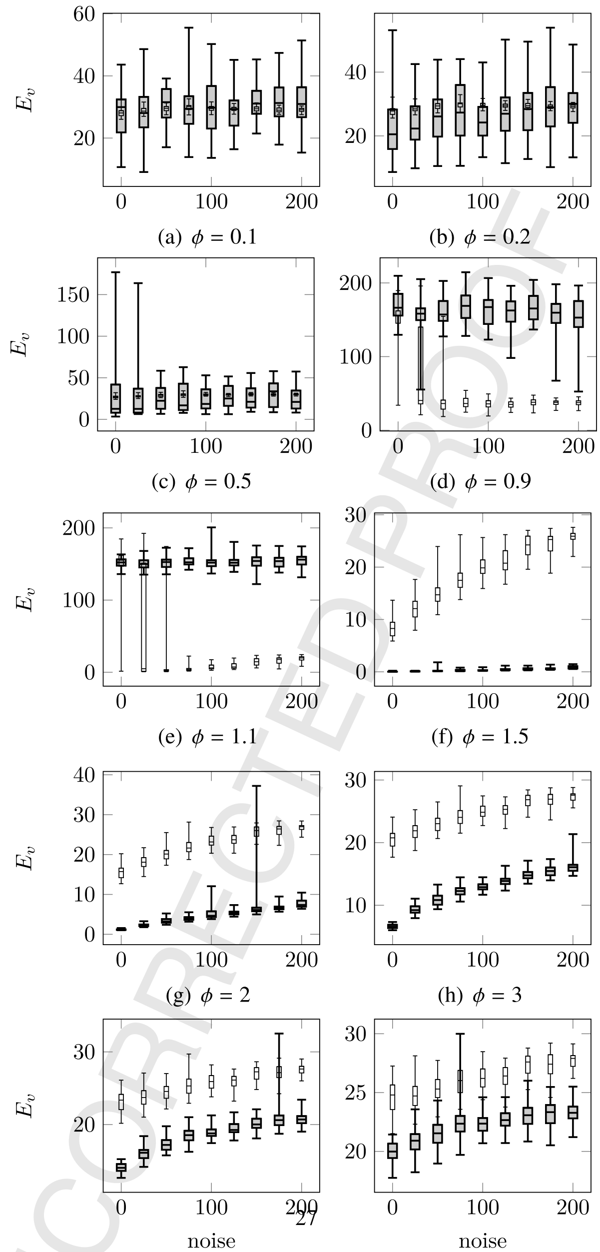



Fig. 14. Changes of elaborated values of attribute weights for the first attribute of the 'art-outliers' data set. Results of five clustering are shown. The first and second clusters are represented by solid and dashed lines. The localisations **V** are initialised data items taken from the data set at random.

Table 8
Medians of error index for weights ($E_z$) for the 'art-noise' dataset elaborated by the FWCOM algorithm with essential loss function.

| cOutliers | Weight parameter $\phi$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0.1 | 0.2 | 0.5 | 1.5 | 2 | 3 | 5 | 10 |
| 0 | 0.3552 | 0.3627 | 0.1574 | 0.5892 | 0.0590 | 0.0213 | 0.0753 | 0.1632 |
| 25 | 0.3354 | 0.3364 | 0.1543 | 1.1398 | 0.3721 | 0.0770 | 0.1101 | 0.1835 |
| 50 | 0.3388 | 0.3394 | 0.3667 | 1.2032 | 0.4332 | 0.1118 | 0.1276 | 0.1954 |
| 75 | 0.3150 | 0.3312 | 0.2087 | 1.2191 | 0.5236 | 0.1398 | 0.1471 | 0.2070 |
| 100 | 0.3424 | 0.3076 | 0.1787 | 1.2195 | 0.5211 | 0.1486 | 0.1574 | 0.2123 |
| 125 | 0.3030 | 0.3156 | 0.9653 | 1.1987 | 0.5039 | 0.1551 | 0.1670 | 0.2173 |
| 150 | 0.3094 | 0.3201 | 0.2724 | 1.1952 | 0.5009 | 0.1603 | 0.1741 | 0.2221 |
| 175 | 0.3073 | 0.3252 | 0.8226 | 1.2026 | 0.4931 | 0.1681 | 0.1824 | 0.2257 |
| 200 | 0.3053 | 0.2986 | 0.3942 | 1.1821 | 0.4446 | 0.1682 | 0.1853 | 0.2285 |

Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001




(a) $\phi = 0.1$ (b) $\phi = 0.2$ (c) $\phi = 0.5$ (d) $\phi = 0.9$ (e) $\phi = 1.1$ (f) $\phi = 1.5$ (g) $\phi = 2$ (h) $\phi = 3$ (i) $\phi = 5$ (j) $\phi = 10$

Fig. 15. The box plot of error index for centre localisation ($E_v$) for the 'art-noise' dataset elaborated by FWCM (narrow boxes) and FWCOM (grey wide boxes) with essential loss function.



The proposed algorithm is an iterative algorithm applying the Picard iteration technique. Figs. 12 and 13 present the convergence of cluster centres and attributes weights in consecutive iterations of the algorithm. Every experiment was repeated 50 times, but only five repetitions are shown in the figures to keep them legible. The figures present values of the first attribute of both clusters (represented by solid and dashed lines) – localisations of the centres of the clusters and the weights of the first attribute in both clusters. The membership matrix **U** and weight matrix **Z** are initialised with random numbers. The centres of clusters are calculated with **U** matrix with Eq. (30). Fig. 14 presents slightly different situation – the cluster centre matrix **V** is initialised with random data items (prototypes) taken from the data set. In clustering procedure the localisations of clusters centres converge to their final values. It is important

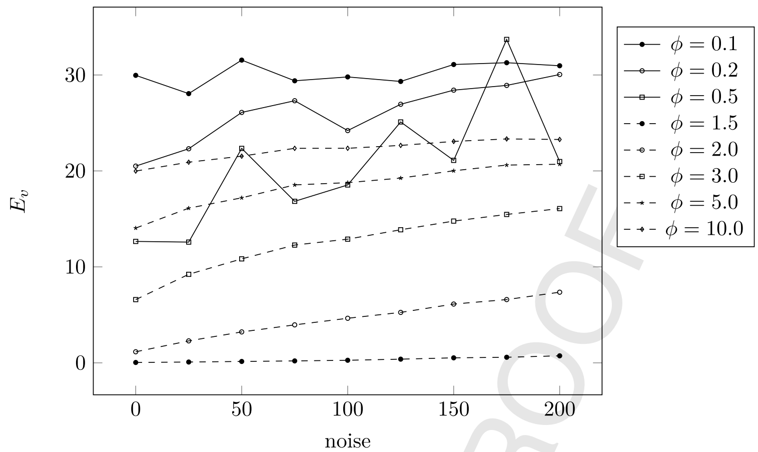



Fig. 16. The median of error index for centre localisation ($E_v$) for the 'art-noise' dataset elaborated by FWCOM algorithm, loss function: essential.

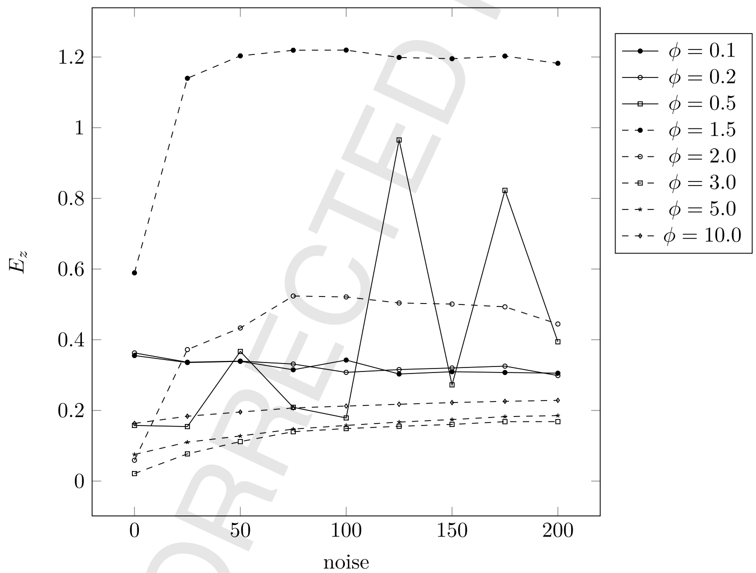



Fig. 17. Medians of error index for weights ($E_z$) for the 'art-noise' dataset elaborated by the FWCOM algorithm with essential loss function.

Table 9
Medians of error index for centre localisation ($E_v$) for the 'art-noise' dataset for weight parameter $\phi = 2$ elaborated with FWCOM algorithm.

| Outliers | Loss function | | | | | | |
|---|---|---|---|---|---|---|---|
| | lin | ess | log | log-lin | sig | sig-lin | Huber |
| 0 | 1.161 | 1.152 | 1.155 | 1.156 | 1.177 | 1.209 | 1.154 |
| 25 | 2.338 | 2.287 | 2.415 | 2.380 | 2.409 | 2.388 | 2.383 |
| 50 | 3.094 | 3.221 | 3.313 | 3.293 | 3.501 | 3.432 | 3.537 |
| 75 | 3.911 | 3.957 | 4.256 | 3.998 | 4.283 | 4.320 | 4.213 |
| 100 | 4.408 | 4.643 | 5.176 | 4.467 | 5.240 | 5.027 | 5.200 |
| 125 | 5.287 | 5.256 | 5.597 | 5.267 | 5.525 | 5.898 | 5.715 |
| 150 | 6.077 | 6.125 | 6.550 | 6.260 | 6.474 | 6.483 | 6.596 |
| 175 | 6.538 | 6.592 | 7.413 | 6.881 | 7.886 | 6.898 | 6.885 |
| 200 | 7.318 | 7.361 | 7.758 | 7.382 | 7.586 | 7.680 | 7.683 |



Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001

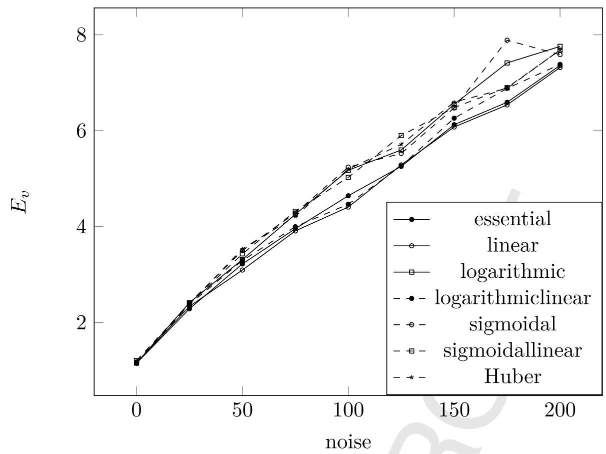



Fig. 18. Medians of error index for centre localisation ($E_v$) for the 'art-noise' dataset for weight parameter $\phi = 2$ elaborated with the FWCOM algorithm.

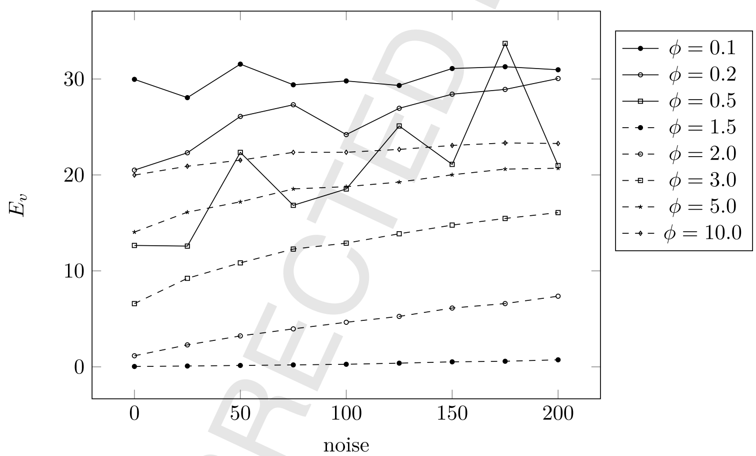



Fig. 19. The median of error index for centre localisation ($E_v$) for the 'art-noise' dataset elaborated by FWCOM algorithm, loss function: essential.

Table 10
The median of error index for centre localisation ($E_v$) for the 'art-noise' dataset elaborated by FWCOM algorithm, loss function: essential.

| Outliers | Weight parameter $\phi$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0.1 | 0.2 | 0.5 | 1.5 | 2 | 3 | 5 | 10 |
| 0 | 29.961 | 20.502 | 12.652 | 0.037 | 1.152 | 6.589 | 14.043 | 19.984 |
| 25 | 28.056 | 22.309 | 12.583 | 0.084 | 2.287 | 9.224 | 16.117 | 20.908 |
| 50 | 31.548 | 26.096 | 22.348 | 0.143 | 3.221 | 10.834 | 17.196 | 21.536 |
| 75 | 29.393 | 27.314 | 16.833 | 0.200 | 3.957 | 12.272 | 18.542 | 22.354 |
| 100 | 29.795 | 24.205 | 18.537 | 0.266 | 4.643 | 12.890 | 18.778 | 22.359 |
| 125 | 29.324 | 26.946 | 25.103 | 0.382 | 5.256 | 13.867 | 19.254 | 22.659 |
| 150 | 31.091 | 28.416 | 21.100 | 0.521 | 6.125 | 14.773 | 20.015 | 23.073 |
| 175 | 31.277 | 28.912 | 33.704 | 0.580 | 6.592 | 15.465 | 20.609 | 23.335 |
| 200 | 30.961 | 30.046 | 20.975 | 0.731 | 7.361 | 16.074 | 20.696 | 23.277 |

to mention that the membership matrix and attribute weight matrix cannot be initialised with only one value. In such a case the centres of all clusters are exactly in the same point and the distances of a data item to the centres of all clusters are the same and the memberships of data items to clusters will not be modified in the iterations of the algorithm.

Table 11
Medians of error index for weights ($E_z$) for the 'art-noise' dataset for weight parameter $\phi = 3$ elaborated with FWCOM algorithm.

| Outliers | Loss function | | | | | | |
|---|---|---|---|---|---|---|---|
| | lin | ess | log | log-lin | sig | sig-lin | Huber |
| 0 | 0.021 | 0.021 | 0.021 | 0.021 | 0.020 | 0.024 | 0.021 |
| 25 | 0.083 | 0.077 | 0.078 | 0.080 | 0.074 | 0.076 | 0.079 |
| 50 | 0.110 | 0.111 | 0.108 | 0.107 | 0.109 | 0.105 | 0.108 |
| 75 | 0.143 | 0.139 | 0.132 | 0.136 | 0.126 | 0.129 | 0.134 |
| 100 | 0.155 | 0.148 | 0.144 | 0.147 | 0.141 | 0.141 | 0.147 |
| 125 | 0.158 | 0.155 | 0.151 | 0.152 | 0.148 | 0.149 | 0.157 |
| 150 | 0.169 | 0.160 | 0.159 | 0.161 | 0.156 | 0.157 | 0.160 |
| 175 | 0.174 | 0.168 | 0.167 | 0.169 | 0.165 | 0.162 | 0.171 |
| 200 | 0.171 | 0.168 | 0.166 | 0.167 | 0.165 | 0.163 | 0.164 |

Table 12
The median of error index for centre localisation ($E_v$) for the 'art-noise' dataset elaborated by FWCOM algorithm for weight parameter $\phi = 1.5$.

| Outliers | Loss function | | | | | | |
|---|---|---|---|---|---|---|---|
| | lin | ess | log | log-lin | sig | sig-lin | Huber |
| 0 | 0.041 | 0.037 | 0.037 | 0.037 | 0.048 | 0.046 | 0.037 |
| 25 | 0.081 | 0.084 | 0.079 | 0.086 | 0.123 | 0.091 | 0.082 |
| 50 | 0.131 | 0.143 | 0.138 | 0.133 | 0.198 | 0.178 | 0.150 |
| 75 | 0.174 | 0.200 | 0.184 | 0.204 | 0.231 | 0.230 | 0.186 |
| 100 | 0.254 | 0.266 | 0.271 | 0.293 | 0.323 | 0.313 | 0.273 |
| 125 | 0.349 | 0.382 | 0.356 | 0.369 | 0.370 | 0.426 | 0.356 |
| 150 | 0.456 | 0.521 | 0.465 | 0.531 | 0.579 | 0.565 | 0.548 |
| 175 | 0.575 | 0.580 | 0.584 | 0.602 | 0.618 | 0.598 | 0.616 |
| 200 | 0.733 | 0.731 | 0.703 | 0.773 | 0.778 | 0.847 | 0.741 |

Table 13
Comparison of errors for centre (prototype) localisation for the 'art-outliers' data set. Abbreviation 'ess' stands for essential loss function, 'sig' – sigmoidal loss function.

| Outliers | Method | | | | | | |
|---|---|---|---|---|---|---|---|
| | FCM | AFGKM | FWCM | | | FWCOM, $\phi = 2$ | |
| | | | $\phi = 1.5$ | $\phi = 2$ | $\phi = 3$ | ess | sig |
| 0 | 8.3250 | 1.45305 | 7.5791 | 14.8517 | 20.0028 | 1.16396 | 1.1365 |
| 5 | 0.6111 | 2.27126 | 6.9952 | 13.8560 | 18.6064 | 1.18659 | 1.1630 |
| 10 | 11.6126 | 2.69218 | 6.2520 | 12.5202 | 16.8278 | 1.23180 | 1.2045 |
| 15 | 0.5665 | 2.95046 | 5.7753 | 11.4588 | 15.4837 | 1.25131 | 1.2614 |
| 20 | 15.8728 | 3.07225 | 5.3071 | 10.3935 | 13.8738 | 1.25307 | 1.2217 |
| 25 | 0.9791 | 3.19034 | 5.4575 | 9.9533 | 13.0721 | 1.27104 | 1.2640 |
| 30 | 2.0228 | 3.27474 | 5.7461 | 9.7164 | 12.2637 | 1.33404 | 1.3209 |
| 35 | 13.4210 | 3.36086 | 6.6735 | 10.0442 | 12.0270 | 1.35385 | 1.3059 |
| 40 | 13.2226 | 3.36420 | 7.7248 | 10.4495 | 12.0379 | 1.34919 | 1.2768 |
| 45 | 12.8445 | 3.47436 | 8.5539 | 10.8552 | 12.2773 | 1.42303 | 1.4362 |
| 50 | 12.9245 | 3.58228 | 9.2702 | 11.2183 | 12.3990 | 1.40922 | 1.3724 |
| 55 | 12.8710 | 3.66130 | 9.3549 | 11.3853 | 12.5928 | 1.41832 | 1.3428 |
| 60 | 5.7560 | 3.90247 | 9.6686 | 11.6232 | 12.7808 | 1.46199 | 1.4094 |
| 65 | 13.2346 | 3.98796 | 10.2297 | 12.1086 | 13.1098 | 1.50398 | 1.4497 |
| 70 | 6.4403 | 4.15924 | 10.3984 | 11.9950 | 13.1237 | 1.61520 | 1.5302 |
| 75 | 6.5867 | 7.95980 | 10.3946 | 12.1506 | 13.1678 | 1.55499 | 1.6562 |
| 80 | 12.9721 | 4.26247 | 10.5875 | 12.2221 | 13.2995 | 1.64367 | 1.5840 |
| 85 | 6.7690 | 4.56014 | 10.9219 | 12.5535 | 13.3038 | 1.72217 | 1.9714 |
| 90 | 7.1698 | 7.68608 | 10.7827 | 12.3024 | 13.2780 | 1.71603 | 1.7564 |
| 95 | 7.2062 | 8.49444 | 11.1182 | 12.6093 | 13.5343 | 1.66295 | 1.9294 |

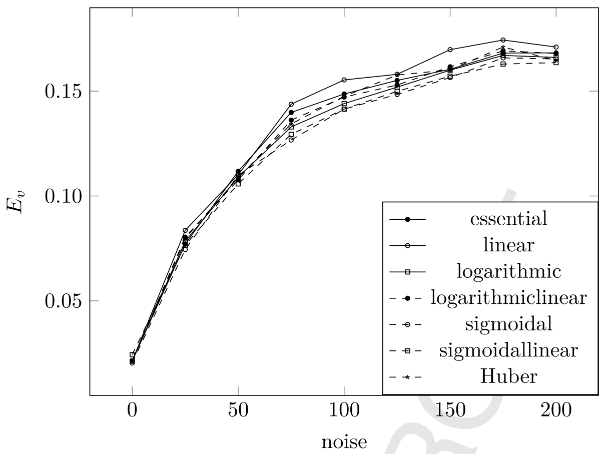



Fig. 20. Medians of error index for weights ($E_z$) for the 'art-noise' dataset for weight parameter $\phi = 3$ elaborated with the FWCOM algorithm.

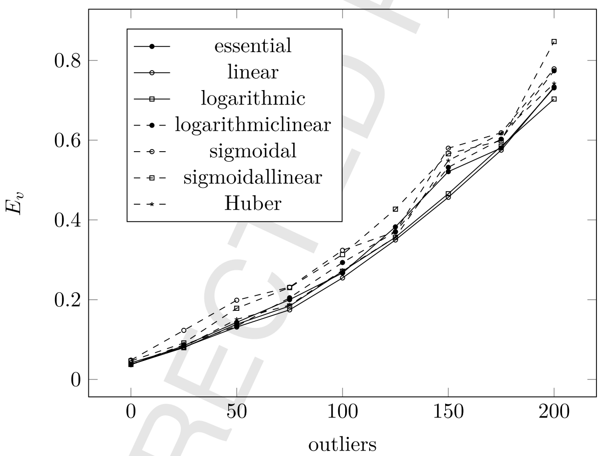



Fig. 21. Medians of error index for centre localisation ($E_v$) for the 'art-noise' dataset for weight parameter $\phi = 1.5$.

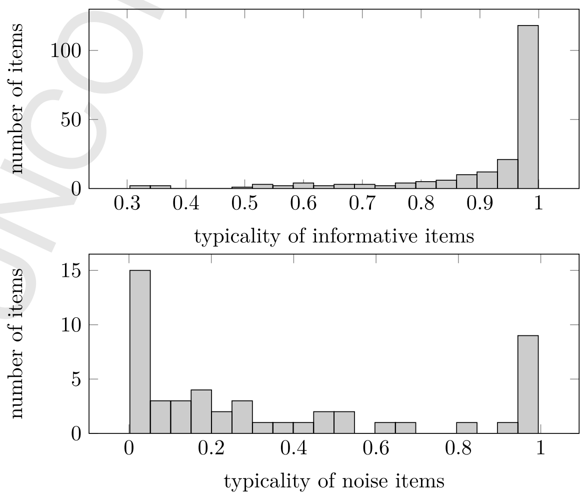



Fig. 22. Histogram of typicalities assigned to informative data items (upper plot) and 50 noise data items (lower plot) by the FWCOM algorithm with essential loss function and $\phi = 2$ ('art-noise' data set).

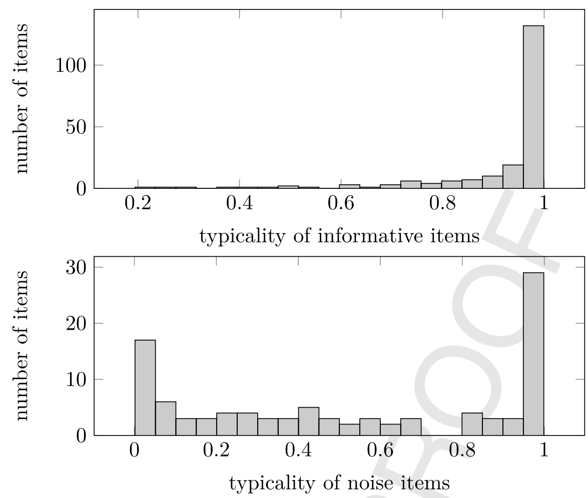



Fig. 23. Histogram of typicalities assigned to informative data items (upper plot) and 100 noise data items (lower plot) by the FWCOM algorithm with essential loss function and $\phi = 2$ (‘art-noise’ data set).

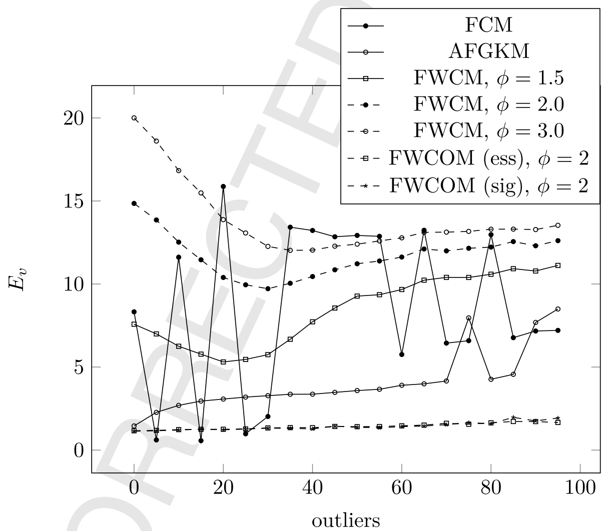



Fig. 24. Comparison of errors for centre (prototype) localisation for the ‘art-outliers’ data set. Abbreviation ‘ess’ stands for essential loss function, ‘sig’ – sigmoidal loss function.

### 3.2.2. *Artificial dataset with background noise: ‘art-noise’*

The data set ‘art-noise’ has the same clusters as the ‘art-outliers’ data set, but instead of outliers it has a background noise. In each cluster there are 100 data items. The number of noise data item is 0, 25, 50, 75, 100, 125, 150, 175, and 200. The values of $\phi$ are the same as in the experiments with the ‘art-outliers’ data set. The number of clusters $C = 2$. Fig. 15 presents the box plots for the results elaborated with the FWCM and FWCOM algorithms for the ‘art-noise’ data set. The lowest values of error measure of centre localisation are elaborated with $\phi = 1.5$ (cf. Fig. 16 and Table 7), whereas the lowest values of error measure for attribute weights are elaborated with $\phi = 3$ (cf. Fig. 17 and Table 17). It can be also seen that for $\phi \in [1.5, 5]$ the results of the FWCOM algorithm have shorter interquartile ranges than those of FWCM. (See also Figs. 18, 19 and Tables 8–13.) Figs. 20 and 21 show that the influence of the loss function on the localisation error and weight error is not very significant. The typicalities assigned to data items have different pattern than for the ‘art-outliers’ data set. In the ‘art-noise’ data set some of the noise items are very close to the centres of informative prototypes and are not outliers. These data items have high typicality. This can be observed in Figs. 22 and 23 for the data set with 50 and 100 noise items respectively. For this data set the typicalities elaborated by the



Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001

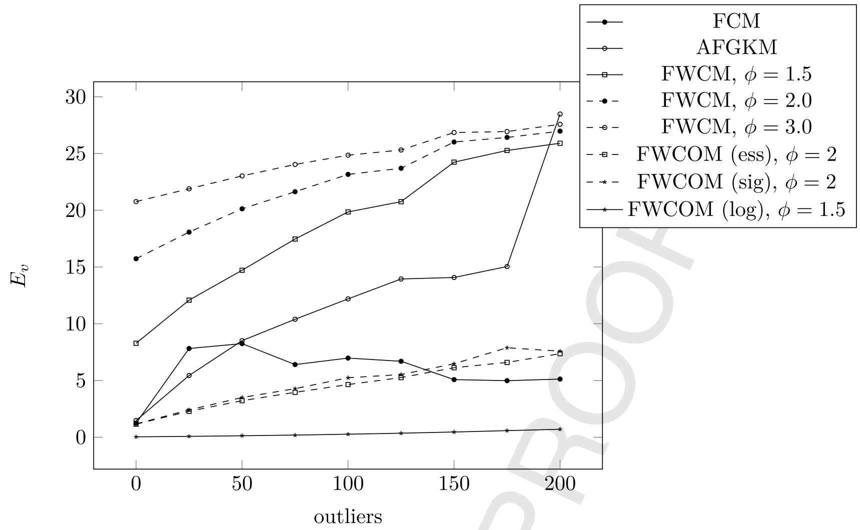



Fig. 25. Comparison of errors for centre (prototype) localisation for the ‘art-noise’ data set. Abbreviation ‘ess’ stands for essential loss function, ‘sig’ – sigmoidal, ‘log’ – logarithmic loss function.

Table 14
Comparison of errors for centre (prototype) localisation for the ‘art-noise’ data set. Abbreviation ‘ess’ stands for essential, ‘sig’ – sigmoidal, ‘hub’ – Huber, ‘lin’ – linear, ‘log’ – logarithmic, ‘lgl’ – logarithmic linear, ‘sig’ – sigmoidal, and ‘sgl’ – sigmoidal linear loss function.

| Method | Number of noise data items | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0 | 25 | 50 | 75 | 100 | 125 | 150 | 175 | 200 |
| FCM | 1.234 | 7.820 | 8.248 | 6.404 | 6.975 | 6.693 | 5.079 | 4.983 | 5.124 |
| AFGKM | 1.477 | 5.439 | 8.508 | 10.394 | 12.191 | 13.944 | 14.075 | 15.040 | 28.476 |
| FWCM | | | | | | | | | |
| $\phi = 1.5$ | 8.280 | 12.078 | 14.713 | 17.466 | 19.851 | 20.754 | 24.237 | 25.268 | 25.902 |
| $\phi = 2$ | 15.728 | 18.073 | 20.125 | 21.633 | 23.158 | 23.698 | 26.011 | 26.416 | 26.968 |
| $\phi = 3$ | 20.761 | 21.892 | 23.024 | 24.033 | 24.850 | 25.307 | 26.848 | 26.932 | 27.574 |
| FWCOM | | | | | | | | | |
| $\phi = 2$, ess | 1.152 | 2.287 | 3.221 | 3.957 | 4.643 | 5.256 | 6.125 | 6.592 | 7.361 |
| $\phi = 2$, sig | 1.177 | 2.409 | 3.501 | 4.283 | 5.240 | 5.525 | 6.474 | 7.886 | 7.586 |
| FWCOM | | | | | | | | | |
| $\phi = 1.5$, hub | 0.037 | 0.082 | 0.150 | 0.186 | 0.273 | 0.356 | 0.548 | 0.616 | 0.741 |
| $\phi = 1.5$, ess | 0.037 | 0.084 | 0.143 | 0.200 | 0.266 | 0.382 | 0.521 | 0.580 | 0.731 |
| $\phi = 1.5$, lin | 0.041 | 0.081 | 0.131 | 0.174 | 0.254 | 0.349 | 0.456 | 0.575 | 0.733 |
| $\phi = 1.5$, log | 0.037 | 0.079 | 0.138 | 0.184 | 0.271 | 0.356 | 0.465 | 0.584 | 0.703 |
| $\phi = 1.5$, lgl | 0.037 | 0.086 | 0.133 | 0.204 | 0.293 | 0.369 | 0.531 | 0.602 | 0.773 |
| $\phi = 1.5$, sig | 0.048 | 0.123 | 0.198 | 0.231 | 0.323 | 0.370 | 0.579 | 0.618 | 0.778 |
| $\phi = 1.5$, sgl | 0.046 | 0.091 | 0.178 | 0.230 | 0.313 | 0.426 | 0.565 | 0.598 | 0.847 |

FWCOM algorithm differ in a significant way from typicalities elaborated for the ‘art-outliers’ dataset (cf. Sec. 3.2.1). It is impossible to name the divisive typicality with noise having lower and informative items higher typicalities. Most of informative items have very high typicalities (similarly to ‘art-outliers’ data set), but the noninformative items (noise) have almost all values of typicality. It is caused by the distribution of noise in the data set. Some noise items lie very close to the cluster centres thus they have high typicality and are treated by the clustering algorithm as typical values. The accuracies defined with Eq. (45) and (46) for data with 50 noise items (Fig. 22) are: for informative data items $\xi_i = 0.8325$ and for noise $\xi_n = 0.4898$. The accuracy for data with 100 noise items (Fig. 23) are: for informative

Table 15
Comparison of medians of clustering time (in [ms]) for the 'art-outliers' data set.

| Outliers | Clustering algorithm | | | | |
|---|---|---|---|---|---|
| | FCM | FCOM | FWCM | FWCOM | AGFKM |
| 0 | 14 | 915 | 43 | 117 | 15 |
| 5 | 14 | 938 | 43 | 162 | 15 |
| 10 | 15 | 982 | 46 | 131 | 16 |
| 15 | 15 | 1291 | 45 | 172 | 16 |
| 20 | 16 | 1357 | 63 | 176 | 17 |
| 25 | 16 | 1052 | 47 | 178 | 16 |
| 30 | 16 | 1422 | 48 | 181 | 17 |
| 35 | 16 | 1464 | 48 | 142 | 18 |
| 40 | 17 | 1298 | 52 | 189 | 18 |
| 45 | 18 | 1332 | 51 | 146 | 18 |
| 50 | 18 | 1629 | 55 | 146 | 20 |
| 55 | 19 | 1601 | 54 | 149 | 19 |
| 60 | 19 | 1593 | 57 | 150 | 20 |
| 65 | 20 | 1235 | 80 | 152 | 20 |
| 70 | 20 | 1755 | 58 | 157 | 20 |
| 75 | 20 | 1832 | 57 | 165 | 21 |
| 80 | 20 | 1514 | 59 | 166 | 21 |
| 85 | 21 | 1595 | 59 | 225 | 21 |
| 90 | 21 | 1894 | 60 | 222 | 21 |
| 95 | 29 | 1404 | 61 | 172 | 23 |
| 100 | 22 | 1488 | 65 | 180 | 22 |

Table 16
Comparison of medians of clustering time (in [ms]) for the 'art-noise' data set.

| Outliers | Clustering algorithm | | | | |
|---|---|---|---|---|---|
| | FCM | FCOM | FWCM | FWCOM | AGFKM |
| 0 | 28 | 289 | 43 | 124 | 15 |
| 25 | 17 | 334 | 49 | 138 | 17 |
| 50 | 18 | 393 | 53 | 157 | 18 |
| 75 | 20 | 435 | 57 | 168 | 20 |
| 100 | 21 | 507 | 63 | 253 | 22 |
| 125 | 24 | 555 | 67 | 205 | 24 |
| 150 | 25 | 828 | 72 | 213 | 25 |
| 175 | 41 | 677 | 78 | 321 | 27 |
| 200 | 28 | 1039 | 83 | 251 | 30 |

Table 17
The weights assigned to attributes in clusters for the 'Methane' data set with the sigmoidal linear loss function. The abbreviation 'attr' stands for 'attribute'.

| attr | Clusters | | | |
|---|---|---|---|---|
| | I | II | III | IV |
| 1 | 0.101 | 0.013 | 0.239 | 0.154 |
| 2 | 0.035 | 0.029 | 0.092 | 0.083 |
| 3 | 0.018 | 0.289 | 0.023 | 0.060 |
| 4 | 0.268 | 0.040 | 0.016 | 0.020 |
| 5 | 0.447 | 0.082 | 0.364 | 0.374 |
| 6 | 0.090 | 0.044 | 0.210 | 0.158 |
| 7 | 0.018 | 0.302 | 0.027 | 0.085 |
| 8 | 0.018 | 0.197 | 0.025 | 0.063 |



Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001

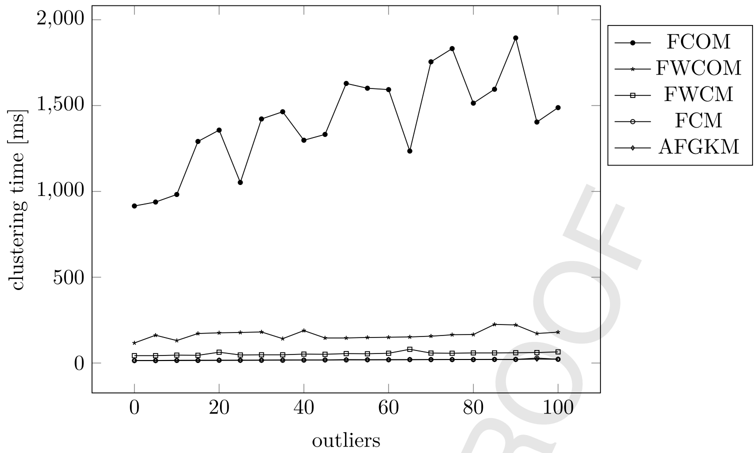



Fig. 26. Comparison of median of clustering time (in [ms]) for the 'art-outliers' data set.

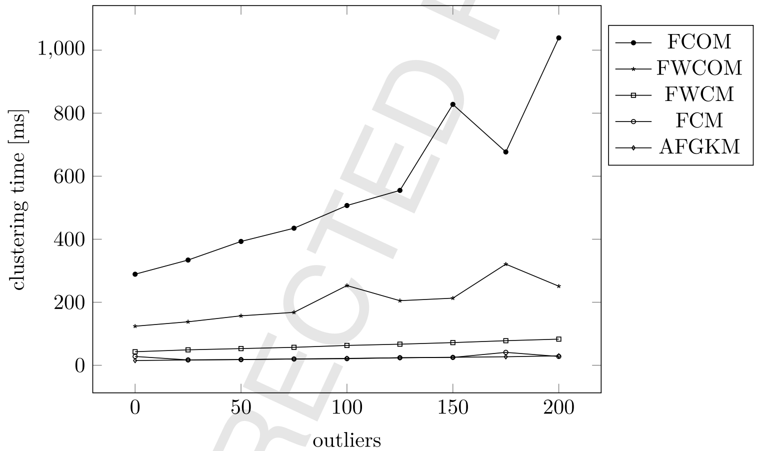



Fig. 27. Comparison of median of clustering time (in [ms]) for the 'art-noise' data set.

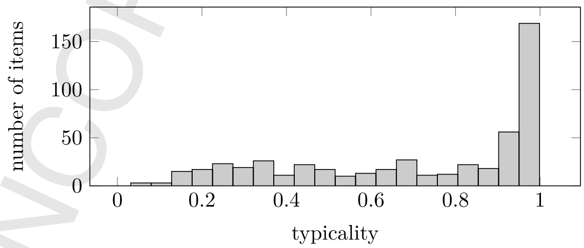



Fig. 28. Histogram of typicalities elaborated by the FWCOM algorithm with sigmoidal loss function for the 'Methane' data set.

data items $\xi_i = 0.8357$ and for noise $\xi_n = 0.3458$. These values show that the ratio of noise has almost no influence on typicalities of informative data items. But the noise ratio has influence on typicalities for noninformative data items. The more noisy the data set is, the greater number of noise items is treated as regular informative data and have high typicalities. It can also observed in Figs. 22 and 23.

### 3.2.3. *Comparison with other clustering methods*

Figs. 24 and 25 present the comparison of the FWCOM algorithm with other clustering techniques: FCM [5], FWCM [19], and AFGKM [9]. For algorithms FWCM and AFGKM, that assign weights to attributes, the same error

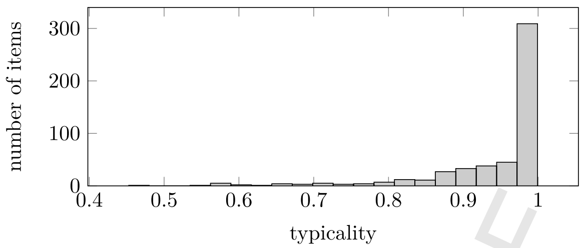



Fig. 29. Histogram of typicalities elaborated by the FWCOM algorithm with linear loss function for the 'Methane' data set.

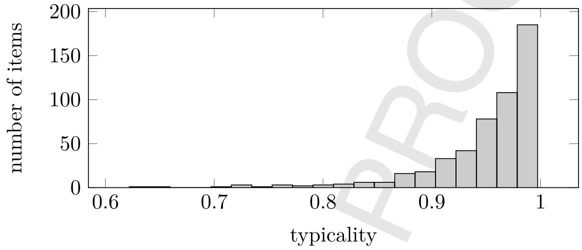



Fig. 30. Histogram of typicalities elaborated by the FWCOM algorithm with Huber loss function for the 'Methane' data set.

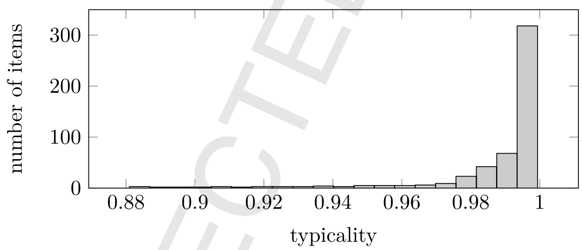



Fig. 31. Histogram of typicalities elaborated by the FWCOM algorithm with essential loss function for the 'Methane' data set.

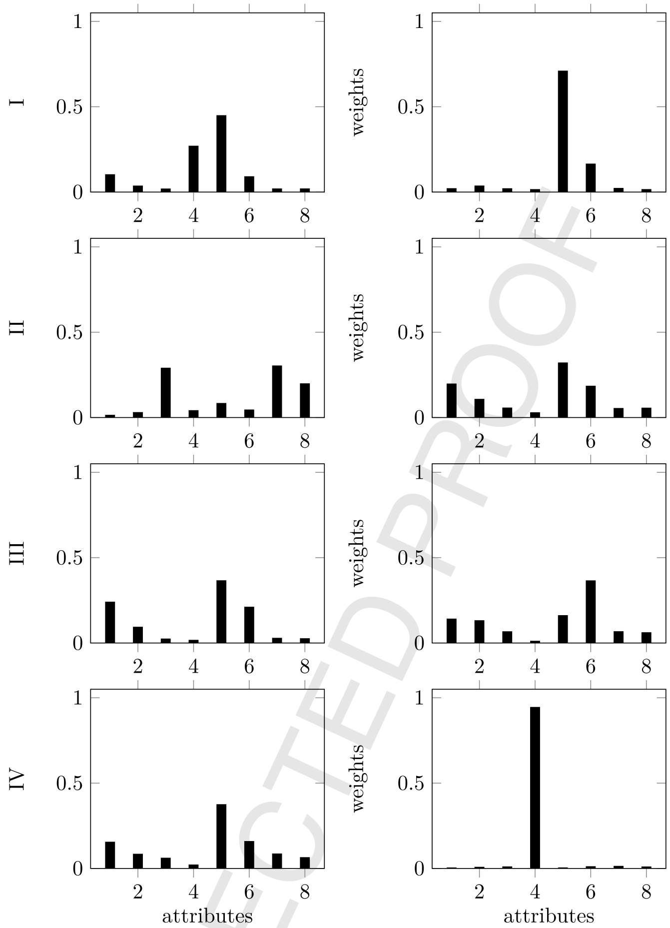
Table 18
The weights assigned to attributes in clusters for the 'Methane' data set with the logarithmic linear loss function. The abbreviation 'attr' stands for 'attribute'.

| attr | Clusters | | | |
|---|---|---|---|---|
| | I | II | III | IV |
| 1 | 0.100 | 0.014 | 0.230 | 0.170 |
| 2 | 0.039 | 0.028 | 0.113 | 0.075 |
| 3 | 0.018 | 0.277 | 0.029 | 0.068 |
| 4 | 0.257 | 0.039 | 0.022 | 0.046 |
| 5 | 0.437 | 0.092 | 0.335 | 0.348 |
| 6 | 0.109 | 0.059 | 0.209 | 0.149 |
| 7 | 0.018 | 0.291 | 0.030 | 0.079 |
| 8 | 0.018 | 0.197 | 0.028 | 0.060 |

measure is used as for the FWCOM, i.e. $E_v$ and $E_z$. For FCM the attributes that should have zero weights are not taken into account. This is why this measure slightly favours FCM and FCOM algorithms, which do not assign weights to attributes.

Fig. 24 presents the error of centre localisation for the 'art-outliers' data set for various clustering algorithms. The lowest errors are elaborated by the FCWOM algorithm. The FCM algorithm elaborated very unstable results. The results elaborated by the FCOM [13] algorithm are significantly higher and are not plotted because they would severely obfuscate other results in the figure.

Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001




(a) sigmoidal linear (b) linear

Fig. 32. Weights assigned to attributes of four clusters for the ‘Methane’ data set by the FWCOM algorithm with two loss functions: sigmoidal linear (left) and linear (right).

Table 19
The weights assigned to attributes in clusters for the ‘Methane’ data set with the Huber loss function. The abbreviation ‘attr’ stands for ‘attribute’.

| attr | Clusters | | | |
|---|---|---|---|---|
| | I | II | III | IV |
| 1 | 0.099 | 0.013 | 0.230 | 0.180 |
| 2 | 0.040 | 0.026 | 0.086 | 0.082 |
| 3 | 0.018 | 0.276 | 0.026 | 0.066 |
| 4 | 0.268 | 0.049 | 0.020 | 0.046 |
| 5 | 0.426 | 0.092 | 0.398 | 0.346 |
| 6 | 0.108 | 0.059 | 0.183 | 0.145 |
| 7 | 0.018 | 0.288 | 0.028 | 0.074 |
| 8 | 0.018 | 0.193 | 0.026 | 0.058 |

The results elaborated for the ‘art-noise’ data set are presented in Table 14. In Fig. 25 for FWCOM, $\phi = 1.5$ only the results for the logarithmic loss functions are plotted. The results elaborated by the FCWOM have the lowest error rates.

The clustering times for the ‘art-outliers’ and ‘art-noise’ are presented in Fig. 26 and Fig. 27 respectively. The number of operations in clustering is independent form both the value of $\phi$ and the applied loss function. In Figs. 26

Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001

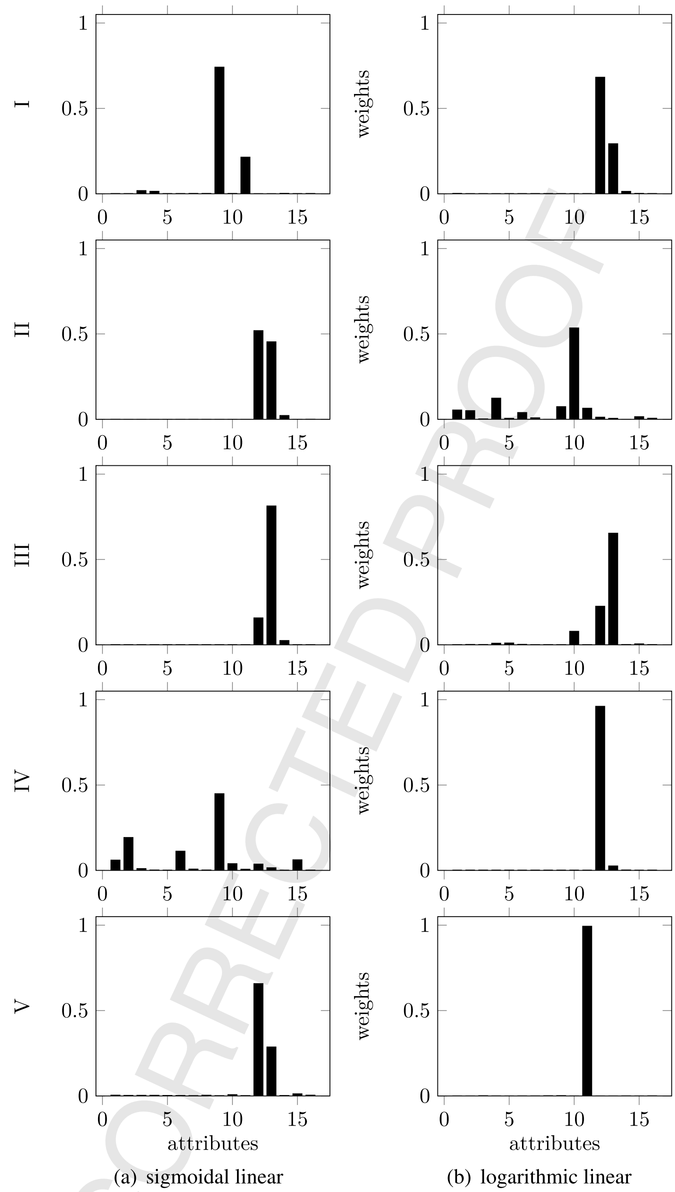



Fig. 33. Weights assigned to attributes of the 'Death' data set by the FWCOM algorithm with two loss functions: sigmoidal linear (left) and logarithmic linear (right).

and 27 we used $\phi = 2$ and 'essential' loss function. Clustering time grows with a number of outliers (noise items) because both outliers and noise items are added to the original data set what results in an increase of the data set size. The fastest algorithms are AFGKM and FCM, the algorithm FWCOM proposed in this paper needs more time than others, but is faster than the FCOM algorithm. The clustering time is a linear function of the data set size, what can be better observed in Tables 15 and 16.

### 3.2.4. Real life datasets

There exist many measures for clustering quality, but it is very hard to find the proper measure to estimate the quality of clustering. In this section we use the real life data sets whose localisations are not known. It is impossible to calculate the quality of precision of localisation of clusters. This is why in this section we will focus on weights assigned to attributes and typicality of data items. In experiments with real life dataset, the parameter $\phi = 2$.

Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001

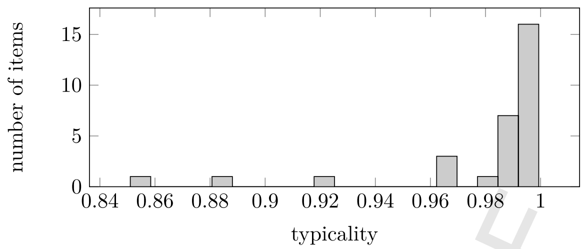



Fig. 34. Histogram of typicalities elaborated by the FWCOM algorithm with linear loss function for the 'Death' data set.

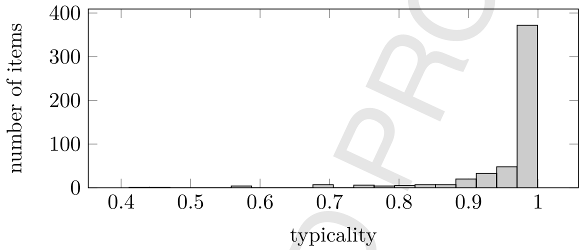



Fig. 35. Histogram of typicalities elaborated by the FWCOM algorithm with essential loss function for the 'Concrete' data set.

Figs. 28, 29, 30, and 31 show the typicalities of data item of the 'Methane' data set elaborated with sigmoidal, linear, Huber, and essential loss functions respectively. In all cases the algorithm elaborated high typicalities for the majority of data items. The lowest typicalities are elaborated with sigmoidal loss function.

Tables 17, 18, and 19 present the weights assigned to attributes in clusters for the 'Methane' data set. The values from Table 17 are plotted in Fig. 32(a). The results need some comment. In the second cluster the 3rd and 8th attributes have high weights. The 8th attributes is a moving sum of 3rd attribute (Sec. 3.1.3). The similar situation can be observed in cluster III. The 1st and 5th, 2nd and 6th attributes build pairs of attributes of similar weights. In cluster VI the attributes representing moving sums have higher weights than the summed attributes themselves. The weights elaborated with the linear loss function (Fig. 32(b)) in cluster II are similar to the weights in clusters III and IV for the sigmoidal linear loss function. Cluster I shows very high weight for the 5th attribute (moving sum of measurements of air flow in a coal mine shaft), cluster II has the highest weight for the 4th cluster (production of coal). This is similar to cluster I for the sigmoidal linear loss function in Fig. 32(a), but here both attributes have high weights in one cluster.

Fig. 33(a) presents the weights assigned to the attributes by the FWCOM algorithm with the sigmoidal linear loss function for the 'Death' data set. In clusters II, III, and V the most important attributes are 12 (hydrocarbon pollution index) and 13 (nitric oxide pollution). In cluster II their significance is similar, in cluster III attribute 13 is assigned higher weight, and in cluster V the most important of those two is attributes 12. In clusters I and IV the most important attribute is 9, but in cluster IV some other attributes are assigned with visible weights. The weights assigned by the FWCOM algorithm with the logarithmic linear loss function are presented in Fig. 33. The attributes with the highest weights are: 12th attribute (hydrocarbon pollution index) in cluster I, 10th (number of office workers) in cluster II, 13th (nitric oxide pollution index) in cluster III, 12th (hydrocarbon pollution index) in cluster IV, and 11th (income lower than $3000) in cluster V. (See also Figs. 34–36.)

Figs. 37(a) and 37(b) present the weights assigned to attributes by the FWCOM algorithm with essential and linear loss functions for the 'Concrete' data set. For both loss functions the most important attributes are: 2 (amount of blast furnace slag), 3 (fly ash), and 8 (fine aggregate). In one cluster the intermediate for all attributes and higher for 7th attribute (coarse aggregate).

The medians of clustering time needed by the FWCOM algorithm are following: for the 'Methane' data set: 843 [ms], 'Death' – 93 [ms], and 'Concrete' – 1078 [ms].



Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001

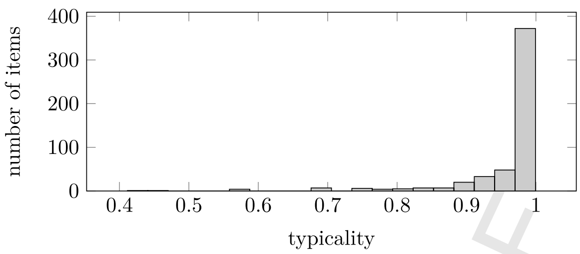



Fig. 36. Histogram of typicalities elaborated by the FWCOM algorithm with linear loss function for the 'Concrete' data set.

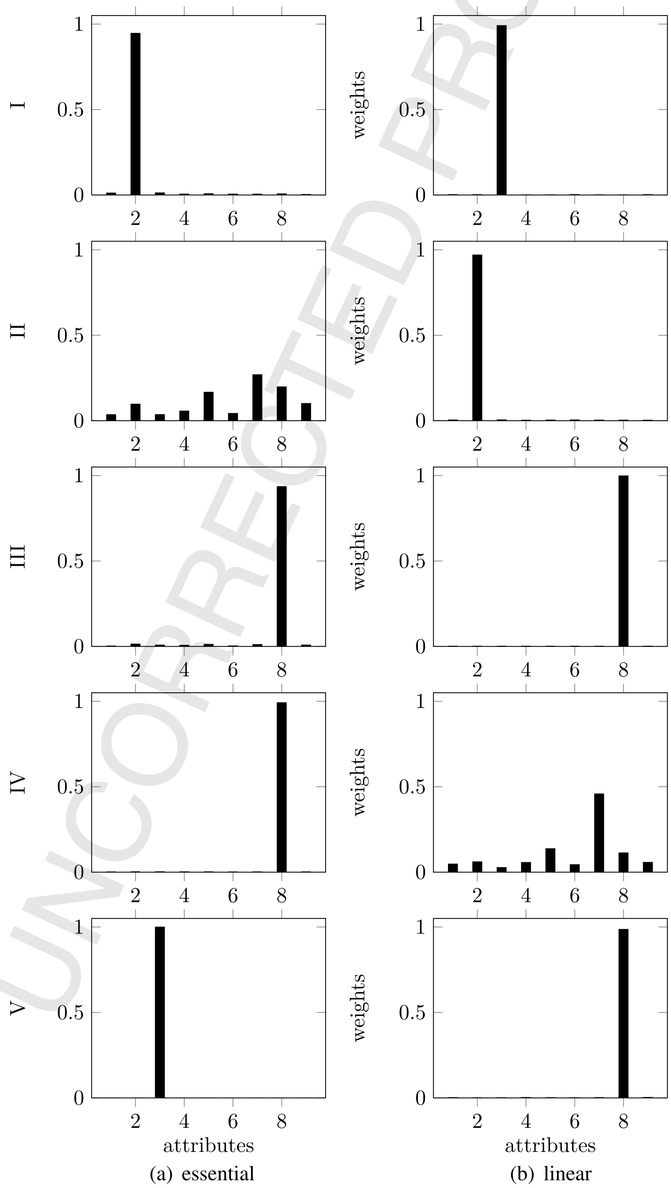



Fig. 37. Weights assigned to attributes of the 'Concrete' data set by the FWCOM algorithm with two loss functions: essential (left) and linear (right).

## 4. Conclusions and future work

The attributes of data sets do not always have the same importance. The importance of attributes may be elaborated with subspace clustering. One of the problems of the soft weight clustering algorithm is their vulnerability to noise and outliers.

This paper presents a fuzzy weighted C-ordered means clustering algorithm that addresses both problems with ordering of data items and attribute weights. The first technique makes the proposed algorithm robust to outliers and noise, the second—elaborates the clusters in subspaces. Thus the paper combines the ordering technique, loss functions, and weighting of attributes separatedly in each cluster. The proposed clustering algorithm is based on a minimalisation of the criterion function. The paper presents the conditions for elaboration of local minimum of the criterion function. The algorithm can handle large ratio of outliers and noise (even up to the cluster cardinality) without severe distortion of both localisations of clusters and weights of attributes. The numerical experiments show that the proposed algorithm can handle data sets with noise or outliers better than compared algorithms.

Clustering is one of the methods used in creation of rule base in neuro-fuzzy systems. The proposed clustering algorithm is robust to noise and outliers and its application in extracting rules from the presented data. This may effectively reduce the vulnerability of fuzzy models to noise and outliers and increase precision of neuro-fuzzy systems for noisy data. The proposed algorithm applies the Picard iterative technique and Lagrangian multipliers. This is a local optimisation approach and thus the algorithm can be unable to leave the local minimum. Some global or global-local techniques may be applied to address this problem.

## Acknowledgements

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

This research was supported by statutory funds (BK-219/RAU2/2016) of the Institute of Informatics, Silesian University of Technology (Gliwice, Poland).

## References

[1] Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu, Cecilia Procopiuc, Jong Soo Park, Fast algorithms for projected clustering, SIGMOD Rec. 28 (2) (1999) 61–72.
[2] Charu C. Aggarwal, Philip S. Yu, Finding generalized projected clusters in high dimensional spaces, in: SIGMOD '00: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, ACM, New York, NY, USA, 2000, pp. 70–81.
[3] Rakesh Agrawal, Johannes Gehrke, Dimitrios Gunopulos, Prabhakar Raghavan, Automatic subspace clustering of high dimensional data for data mining applications, SIGMOD Rec. 27 (2) (1998) 94–105.
[4] Chun-Hung Cheng, Ada Waichee Fu, Yi Zhang, Entropy-based subspace clustering for mining numerical data, in: KDD '99: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY, USA, 1999, pp. 84–93.
[5] J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact, well separated clusters, J. Cybern. 3 (3) (1973) 32–57.
[6] A. Frank, A. Asuncion, UCI machine learning repository, 2010.
[7] J.H. Friedman, J.J. Meulman, Clustering objects on subsets of attributes, J. R. Stat. Soc. B 66 (2004) 815–849.
[8] G. Gan, J. Wu, A convergence theorem for the fuzzy subspace clustering (FSC) algorithm, Pattern Recognit. 41 (6) (2008) 1939–1947.
[9] Guojun Gan, Michael Kwok-Po Ng, Subspace clustering with automatic feature grouping, Pattern Recognit. (2015).
[10] Guojun Gan, Jianhong Wu, Zijiang Yang, A fuzzy subspace algorithm for clustering high dimensional data, in: Advanced Data Mining and Applications, Second International Conference, ADMA 2006, Xi'an, China, August 14–16, 2006, Proceedings, in: Lecture Notes in Computer Science, vol. 4093, Springer, Berlin/Heidelberg, 2006, pp. 271–278.
[11] Sanjay Goil, Sanjay Goil, Harsha Nagesh, Harsha Nagesh, Alok Choudhary, Alok Choudhary, Mafia: efficient and scalable subspace clustering for very large data sets, Technical report, 1999.
[12] Jacek Leski, Marian Kotas, On robust fuzzy $c$-regression models, Fuzzy Sets Syst. 279 (2015) 112–129.
[13] Jacek M. Leski, Fuzzy $c$-ordered-means clustering, Fuzzy Sets Syst. (2014).
[14] Lance Parsons, Ehtesham Haque, Huan Liu, Subspace clustering for high dimensional data: a review, ACM SIGKDD Explor. Newsl. 6 (1) (2004) 90–105.
[15] Leszek Rutkowski, Krzysztof Cpałka, Flexible neuro-fuzzy systems, IEEE Trans. Neural Netw. 14 (3) (May 2003) 554–574.
[16] Marek Sikora, Beata Sikora, Application of machine learning for prediction a methane concentration in a coal-mine, Arch. Min. Sci. 51 (4) (2006) 475–492.
[17] Krzysztof Siminski, Clustering in fuzzy subspaces, Theor. Appl. Informatics 24 (4) (2012) 313–326.
[18] Krzysztof Siminski, Clustering with missing values, Fundam. Inform. 123 (3) (2013) 331–350.
[19] Krzysztof Siminski, Neuro-fuzzy system with weighted attributes, Soft Comput. 18 (2) (2014) 285–297.

[20] Krzysztof Siminski, Rough subspace neuro-fuzzy system, Fuzzy Sets Syst. 269 (2015) 30–46.
[21] Helmuth Späth, Mathematical Algorithms for Linear Regression, Academic Press Professional, Inc., San Diego, CA, USA, 1992.
[22] Jiong Yang, Wei Wang, Haixun Wang, P. Yu, $\delta$-Clusters: capturing subspace correlation in a large data set, in: Proceedings. 18th International Conference on Data Engineering, 2002, pp. 517–528.
[23] I. Cheng Yeh, Modeling of strength of high-performance concrete using artificial neural networks, Cem. Concr. Res. 28 (12) (1998) 1797–1808.



Please cite this article in press as: K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets Syst. (2017), http://dx.doi.org/10.1016/j.fss.2017.01.001