

A Rapid Soft Computing Approach to Dimensionality Reduction in Model Construction

Vesa A. Niskanen

Dept. of Economics & Management,
PO Box 27, University of Helsinki, 00014 Helsinki, Finland
vesa.a.niskanen@helsinki.fi

1. Introduction

In model construction large observation or input variable sets arouse various problems and thus we usually attempt to simplify our examinations by applying data compression or dimensionality reduction. Then we may operate with fewer observations or variables. In statistical multivariate analysis this means that in the former case we may compress our data matrices by applying cluster analysis (CA), whereas the number of variables is reduced by combining similar original variables for variable groups with such methods as the principal component analysis (PCA) and factor analysis (FA) [13].

Today soft computing (SC) systems have also proven to be useful in statistical modeling and model construction in general, and its CA and regression models provide good examples of these [1,11,28]. On the other hand, we still face certain challenges when applying SC techniques to dimensionality reduction.

One open problem is how to reduce the dimensionality in our data matrix if the traditional PCA and FA approaches are insufficient. Typical limitations of PCA and FA are that they are only appropriate for linear models, their data sets should be sufficiently large and their variables are expected to be normally distributed [13,39]. In many practical cases, however, actual relationships are nonlinear and/or sample sizes are small. In such situations, SC techniques have often been very successful in data processing. It is therefore reasonable to believe that SC techniques will be also efficient in reducing dimensionality

Below we apply SC models to dimensionality reduction and we aim at a simple and an easily understandable method that is also robust and good in practice. This

approach thus provides us with a “quick-and-dirty” method in the practice of model construction. We also attempt to draw an analogy between our approach and the PCA and FA. Some SC approaches to dimensionality reduction are available already, but they seem to be fuzzified versions of PCA or FA [4,7,9,10,16,17,18, 21, 22,24,27,31,33].

On the other hand, some papers have applied fuzzy similarity measures to this problem area, but they do not seem to correspond with the theoretical background or the goodness criteria of PCA and FA [21,22]. Our approach, in turn, applies fuzzy similarity measures but we also yield and assess our outcomes according to PCA and FA and their goodness criteria. We also maintain Lotfi Zadeh’s original idea on fuzzy systems, viz. instead of only using fuzzy mathematics or set theory, we also apply fuzzy reasoning in an understandable manner [36,37,38]. Thanks for the good available fuzzy methods in CA and approximate reasoning, that are well-known in the fuzzy community already, we adopt a general, a meta-level, approach, and thus detailed calculations are precluded.

Section 2 presents basic ideas on PCA and FA. Section 3 introduces our method. Section 4 provides two real-world examples. Section 5 concludes our examination.

2. Dimensionality Reduction in Statistics

In the explorative studies in human sciences [23] we aim at reducing the number of the original variables by grouping first the similar variables and then specifying such new variables that constitute these variable groups. These new variables are often referred to as sum variables because they are usually the (possibly weighted) sums the original variables. In this manner we may understand better the nature of our data and can perform simpler calculations. We may even attempt to find new “latent” variables behind the original variables in which case we can also label these new variables according to our interpretations, if necessary. For example, if we notice that a certain group of variables in our data matrix actually measures the same general feature, such as person’s mathematical skills, from various standpoints, we may specify the sum variable “mathematical skills” that captures the meanings of the corresponding original variables. In the confirmatory studies, in turn, we may apply the available background theories to our dimensionality reduction.

In the traditional statistics we may specify our sum variables directly on our intuitive basis by calculating the sums of the selected original variables. We may also apply PCA or FA, in which case the obtained sum variables are referred to as principal components or factors, respectively. In fact, these methods operate with

the standard scores of the original variables, i.e., for each variable its mean is subtracted from the observations, and then these differences are divided by its standard deviation, $\text{standard score} = (\text{observation} - \text{mean}) / \text{standard deviation}$. This transformation indicates us how the observations are distributed around the mean in the units of the standard deviation. Hence, when in PCA and FA we calculate the sum variable values for the original observations, viz. the principal component or factor scores, we also obtain the standard scores, and this may complicate the interpretation of our outcomes. Thus simpler sum variable specifications are also available, and these are discussed below.

We usually proceed as follows with the traditional PCA and FA [13]:

1. We assume that there are sufficiently high linear inter-correlations between the original variables, because the similarities between the variables are usually based on these correlations.
2. We also assume that the sample size is sufficiently large (e.g., at least five observations per variable), outliers are excluded, and the variables are measured at least at the level of good ordinal scales. In FA the variables should also be normally distributed and multicollinearity is not accepted.
3. We calculate the so-called principal component or factor loadings that are the correlations between the original variables and the principal components or factors.
4. We apply rotation to these components or factors in order to better interpret our loadings.
5. We select the appropriate principal components or factors, and these will be used in our sum variable specifications.

One distinction between the PCA and FA is that in the former case we always obtain the same unique components, whereas in FA the factors may vary according to the established number of factors.

In a sense, PCA and FA aim to find variable clusters according to the linear inter-correlations between the variables. The variable groups having inter-correlations will constitute variable clusters, and thus the obtained components or factors are the corresponding cluster centers. The principal component or factor loadings may now be regarded as being the “degrees of membership” of the variables to these clusters. Another approach is to consider that the principal components or factors span such vector spaces in which the loadings of the variables denote their coordinates.

On the other hand, we may also assume that in dimensionality reduction we aim to find clusters of variables according to their distances, and thus we can apply CA or multidimensional scaling. Hence, instead of the so-called Q-techniques of clustering, we are now applying the R-techniques [29]. In practice, when our data set is not too large, we may then operate with the transposed version of the original data matrix and then apply CA, and this approach is adopted below. We also apply fuzzy rule-based reasoning in our analyses. Thanks for the good fuzzy

clustering techniques and usable approximate reasoning methods, the SC approach is more robust, applicable and user-friendly than the traditional methods.

3. Soft Computing and Dimensionality Reduction

Our SC approach below applies both fuzzy clustering and approximate reasoning directly to dimensionality reduction. In this manner we may also apply dimensionality reduction in a nonparametric manner to nonlinear data sets. Since we adopted a meta-level approach, i.e., only methods of general nature are considered and detailed mathematical analyses are precluded, we apply the prevailing fuzzy clustering methods and fuzzy rule-based models, and these have also proved to be useful in practice. However, for the sake of consistency, our goodness criteria for the outputs are those of PCA and FA. In this respect fuzzy mountain clustering [5,32] and fuzzy c-means clustering methods [2,3,6,8,12,15,16,17,20,25,26] are analogous to PCA and FA, respectively.

For example, in mountain clustering and PCA we specify the unique outputs iteratively one at a time starting from the densest or largest group of observations or variables, whereas within the fuzzy c-means method and FA our outcomes vary according to the established number of clusters or factors. We focus on the mountain clustering method and PCA below, because these techniques may bring better understanding to our approach.

Given now the original data matrix with m cases or observations (rows) and n variables (columns), if we apply fuzzy clustering method to dimensionality reduction, we may proceed as follows (Table 3.1):

1. We focus on such groups of variables that are close to each other. In other words, the distances between these variables, X_i , are small. In practice, we may operate, for example, with their standard scores, ZX_i , and our task stems from the calculation of the norms, $\|ZX_i - ZX_j\|$ ($i \neq j$). Alternative transformations may also be used, but in any case the original variables should be transformed into similar scales, because otherwise our variables have unequal weights.
2. Our method uses the transpose of the original data matrix (columns become rows), and then we apply fuzzy cluster analysis to the variables.
3. The obtained cluster centers of the variables, C_k , will be our “principal components” or “factors”, and our loadings are now the correlations between the variables and these centers. We may also use linguistic values in this context.
4. We assess the goodness of our outcomes by applying such prevailing criteria as the communalities and eigenvalues of the variables.

Table 3.1. Example of a data matrix and its principal components.

Case nr.	Variables and principal components					
	X_1	$X_2 \dots$	X_n	C_1	$C_2 \dots$	$C_{q \leq n}$
1	x_{11}	x_{12}	x_{1n}	c_{11}	c_{12}	c_{1q}
2	x_{21}	x_{22}	x_{2n}	c_{21}	c_{22}	c_{2q}
...		
m	x_{m1}	x_{m2}	x_{mn}	c_{m1}	c_{m2}	c_{mq}

When we specify the sum variables according to our loadings, we use functions that express sufficiently well the relationships between the variables and the cluster centers. Traditionally we may apply linear regression analysis in order to obtain such corresponding sum variables, $S_k \approx C_k$,

$$S_k = \sum_i w_{ik} \cdot ZX_i, (i=1, 2, \dots, n), \quad (3.1)$$

in which C_k is the dependent variable, the weights, w , are the regression coefficients and ZX are the standard scores of the original variables [13,14,19,35,39]. This method is used for calculating the principal component or factor scores.

However, (3.1) also includes the irrelevant variables in the sum variables, i.e., the variables with low loadings, and thus it may yield more or less misleading outcomes. Hence, in practice, we quite often simply calculate the sum of the relevant variables in each component, i.e., we only select the variables with the high loadings. In the case of the similar original scales we prefer the averages of these variables. This idea is widely used in PCA and FA in the human sciences. We may then justify our decisions by applying item analysis with Cronbach's alpha reliability coefficients, which are based on correlation coefficients, even though this method is not foolproof for this task [13,39]. Another, sometimes more reliable method is based on the factor score covariance matrix [13]. The former usually minimizes and the latter maximizes the reliability coefficients.

Within our SC framework, we may also apply the fuzzy rule-based models, F , for all relevant variables in a component, i.e., the cluster center, C_k , is the dependent variable in the model

$$S_k = F_k(X_q, \dots, X_r), q \geq 1, r \leq n, \quad (3.2)$$

Method (3.2) seems better in practice because it is also appropriate for nonlinear relationships.

Our SC approach may nevertheless arouse problems if variable clusters are unavailable or we have large observation sets. In the former case it may be difficult

to find plausible cluster centers with any method, and the latter case may lead to quite heavy computations due to the large number of parameters.

Below we apply the idea on mountain clustering that is analogous to PCA [5,32]. Hence, our method specifies the first cluster center according to the greatest or densest variable cluster. The second center is assigned to the second densest cluster with the restriction that it is not in the neighborhood of the first cluster. The next center, in turn, belongs to the third densest cluster, but it should also locate far from the previous clusters, and so forth. The number of cluster centers is determined by the user in the manner of PCA.

From the mathematical standpoint, the general idea for our clustering is that our first cluster center, C_1 , is obtained when we minimize this type of penalty function,

$$\sum_i \mu_{C_1}(ZX_i) \cdot \|ZX_i - C_1\|, i = 1, 2, \dots, n, \quad (3.3)$$

in which μ is an appropriate fuzzy triangular or bell-shaped membership function with its maximum value at C_1 . This method finds the vector C_1 to be the center of the densest variable cluster.

The second cluster center, C_2 , should represent the second densest cluster, and thus the neighborhood of C_1 should be excluded from our analysis. Hence, our minimizing penalty function should now also contain the exclusion function, Ex , that excludes the first cluster,

$$\sum_i Ex_{C_2}(ZX_i) \cdot \mu_{C_2}(ZX_i) \cdot \|ZX_i - C_2\|, i = 1, 2, \dots, n, \quad (3.4)$$

in which, for example, $Ex_{C_2}(ZX_i) = (1 - \mu_{C_1}(ZX_i))^s$ ($s > 1$). This means that in the second round the variables close to C_1 are irrelevant.

In the third round, the variables close to C_1 and C_2 are irrelevant, i.e.,

$$Ex_{C_3}(ZX_i) = \min(Ex_{C_2}(ZX_i), (1 - \mu_{C_2}(ZX_i))^s) \quad (3.5)$$

and the penalty function for C_3 is

$$\sum_j (Ex_{C_3}(ZX_i) \cdot \mu_{C_3}(ZX_i) \cdot \|ZX_i - C_3\|), i = 1, 2, \dots, n. \quad (3.6)$$

We will continue till all the values of $Ex_{C_i}(ZX_i)$ are small, this meaning that we have examined all variable clusters.

These operations may be carried out conveniently with such methods as the genetic algorithms if custom-made models are preferred. We may also apply the original mountain clustering method directly, if the number of variables is sufficiently large. Below we will provide examples with the empiric data sets.

4. Real-World Examples

We examine below two real-world data sets with both PCA and our SC method. We aim to demonstrate that our method, that is simpler and more robust than PCA and FA, is also plausible for dimensionality reduction. We use Matlab™ version 2014b and IBM SPSS™ version 22 in our calculations.

4.1. The Iris Data

Fisher's Iris data is the widely-used benchmark data in cluster analysis. Fuzzy clustering methods have already proved their usability in this context of grouping the objects, but only some indirect methods have been suggested for dimensionality reduction of variables. This data set is challenging to us because it contains problematic clusters.

The Iris data contains 150 of these flowers and four feature variables that measure in millimeters their *Sepal lengths* (Sl), *Sepal widths* (Sw), *Petal lengths* (Pl) and *Petal widths* (Pw; Fig. 4.1). In the cluster analysis for these flowers we should find three clusters, and good fuzzy CA methods are able to perform this. Since sufficiently high inter-correlations prevail between the variables, we may also attempt to use PCA and our SC method for dimensionality reduction.

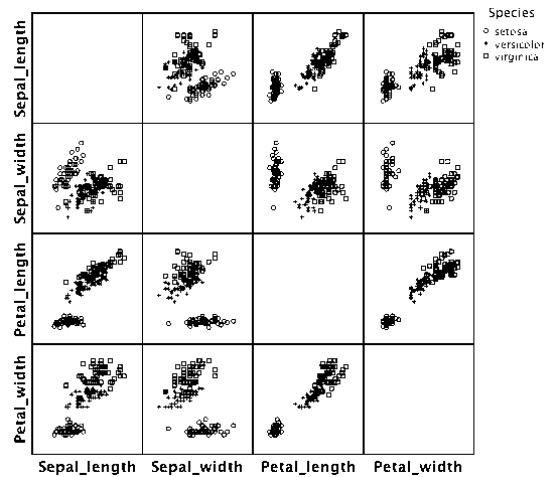


Fig. 4.1. Scatter plots of flowers in the Iris data.

4.1.1. The PCA Approach

In PCA the values of the inter-correlations between our variables may first be analyzed with such rules of thumb as the Kaiser-Mayer-Olkin measure and Bartlett's test, and if the former yields values greater than .6 and the latter rejects its null hypothesis, our correlations seem to be sufficiently high [13,39]. In our data the former yields .54 and the latter rejects the null hypothesis (at the level of significance $< .05$). Hence, the former value is not fully satisfactory, but the latter fulfills the conditions.

On the other hand, the communalities of the variables are higher than .9, and hence we assume that PCA is justified with all our feature variables in this context. The communalities are the *rsquares* in those regression models in which the feature variable is the dependent variable and the principal components are the independent variables. Hence, the communalities indicate how well the selected principal components can explain or predict the variances of the variables.

PCA yields first at the extraction stage the initial component loadings (i.e., the correlations between the variables and the components) for the variables in each component by starting from the largest variable group. In the first component we thus obtain the highest absolute values of the loadings, the second component has the second highest values, and so forth. There is also the restriction that the components must be orthogonal, i.e., they have no inter-correlations.

We also calculate the sums of squares of these loadings in each component, and these sums are referred to as the eigenvalues. We are usually interested in those principal components that yield eigenvalues greater than or equal with unity. The sums of the eigenvalues of our components divided by the number of the original variables, in turn, reveals us how much our components explain of the total variance of our variables.

In order to better understand our outputs, rotation is also carried out, and it modifies our original principal component loadings. The rotation aims to yield either high or low loadings, and in addition to the orthogonal methods, we may now apply oblique methods. The latter methods allow inter-correlations between the principal components, and this situation is usual in the human sciences. However, unlike in orthogonal rotation, in oblique rotation the loadings are not the correlations between the components and variables, but rather the weights that show us the importance of the variables in each component. Typical examples of orthogonal and oblique methods are Varimax and Promax, respectively.

Summing up the foregoing measures, given a table of original or orthogonally rotated principal component loadings, the row-wise sums of squares of the load-

ings yield the communalities, whereas the corresponding column-wise values yield the eigenvalues.

We prefer the oblique “Procrustean” Promax method in rotation below, and these principal component loadings are presented in Table 4.1.1.1 and Fig. 4.1.1.1 (the loadings less than the absolute value of .3 are omitted below because they are irrelevant). We select two principal components, because they already explain approximately 96% of the total variance of the variables (i.e., the sum of these two eigenvalues / 4 · 100% = 96%) even though the eigenvalue of the second component was slightly less than unity.

We notice that, according to the first component in our rotated table, we may generate a sum variable that includes the variables *Sepal length*, *Petal length* and *Petal width*. The second component only includes one high loading, viz. for *Sepal width*. Hence, instead of the original variables, we may use two principal variables within the Iris data, if necessary. Table 4.1.1.2 presents the loadings that are also the correlations between the variables and the components because in oblique rotation the loadings in Table 4.1.1.1 are not correlations (as in the orthogonal rotation). The latter loadings are better comparable with our SC analyses below. We notice that these loadings are slightly more blurred with respect to sum variable specification. Both of these loading tables are nevertheless used in the conduct of inquiry.

Table 4.1.1.1. Rotated component matrix of Iris data

	Component	
	1	2
Sepal_length	1.000	
Petal_length	.933	
Petal_width	.929	
Sepal_width		1.000

Extraction Method: Principal Component Analysis.

Rotation Method: Promax with Kaiser Normalization.

Rotation converged in 3 iterations.

Table 4.1.1.2. The intercorrelations between the variables and the components in Iris data.

	Component	
	1	2
Petal_length	.982	-.455
Petal_width	.962	-.407
Sepal_length	.938	
Sepal_width	-.307	.995

Extraction Method: Principal Component Analysis.
 Rotation Method: Promax with Kaiser Normalization.

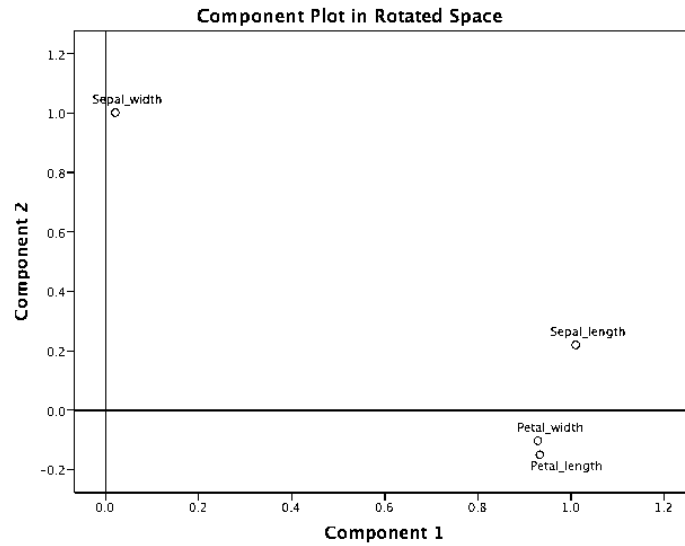


Fig. 4.1.1.1. Component plot in the rotated space with the Iris data.

According to Table 4.1.1.3, that presents the regression coefficients for the component scores, our first sum variable would now be

$$S_1 = .366 \cdot ZS1 + .005 \cdot ZSw + .339 \cdot ZPl + .338 \cdot ZPw \quad (4.1.1.1)$$

if this prevailing method is applied to the standardized feature variables. Since *Sepal width* is irrelevant to S_1 and the rest of the loadings are quite similar, in practice we may use for the original variables their nonweighted sum instead,

$$S_1 = S1 + P1 + Pw, \quad (4.1.1.2)$$

or their average, if their standard scores are used. In item analysis Cronbach's alpha is greater than .9 for S_1 , and this result also corresponds to this sum variable construction.

Table 4.1.1.3. Component Score Coefficient Matrix for Iris data.

	Component	
	1	2
Sepal_length	.366	.200
Sepal_width	.005	.923
Petal_length	.339	-.140
Petal_width	.338	-.097

Extraction Method: Principal Component Analysis.
Rotation Method: Promax with Kaiser Normalization.

Hence, PCA provided us with one plausible sum variable, and this was due to the high linear intercorrelations between the feature variables.

4.1.2. The Soft Computing Approach

If we apply our SC method, we principally utilize the distances between the variables and, in the manner of the PCA, we operate with the standard scores of the original variables. Then, within the Iris data, we notice in the dendrogram in Fig. 4.1.2.1 that Sepal width is clearly distinct from the others and the rest of the feature variables seem to belong to same cluster. The multidimensional scaling analysis (SPSS Proxcal), that allocates the variables into a 2-D space according to their distances, also seems to support quite well this resolution (Fig. 4.1.2.2). Hence, it seems that we may specify one sum variable as above.

According to our cluster analysis approach, we will proceed as follows:

1. We specify two cluster centers, and these are our principal components, C_i . The correlations between the variables and principal components will be our component loadings.
2. The communalities are the *rsquares* of the fuzzy models $F_i: (C_1, C_2) \rightarrow ZX_i$, $i = 1, 2, \dots, n$, i.e., we consider how well our components explain or predict the variables.
3. The eigenvalues are the squared column-wise sums of the loadings as above.

4. The final sum variable, S_1 , only constitutes the relevant variables of the first principal component, and its specification is similar to that of the PCA method. We may apply the fuzzy model $S_1 = F_1(ZX_1, ZX_3, ZX_4)$ by using C_1 as the dependent variable instead, if necessary.

Our mountain clustering method for the feature variables seems to yield two plausible cluster centers, and the corresponding principal components contain sufficiently high loadings. We used Matlab's Fuzzy Logic Toolbox and Takagi-Sugeno reasoning for these tasks [30]. Table 4.1.2.1 presents these correlations, or loadings (the absolute values less than .3 are omitted as above). This Table is analogous to Table 4.1.1.2 within the PCA. For the illustrative purposes, we also calculated the corresponding *rmse* values, and naturally they were consistent with our loadings. (Table 4.1.2.2).

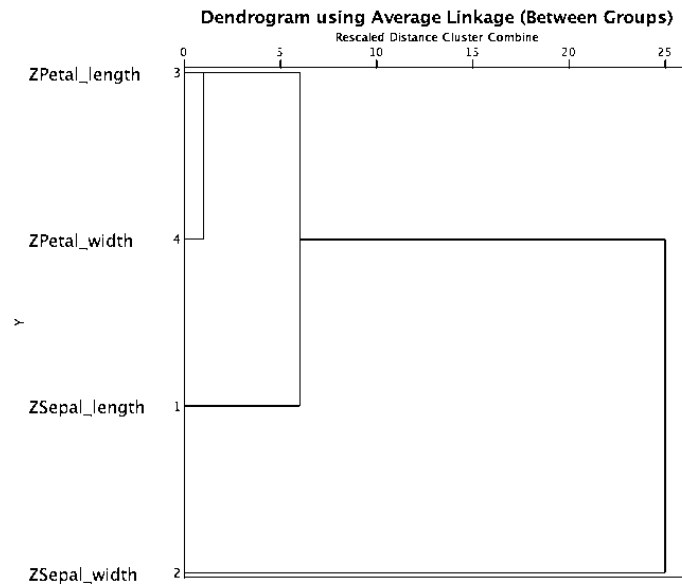


Fig. 4.1.2.1. Dendrogram based on the average linkage method and distances between four standard score variables in the Iris data.

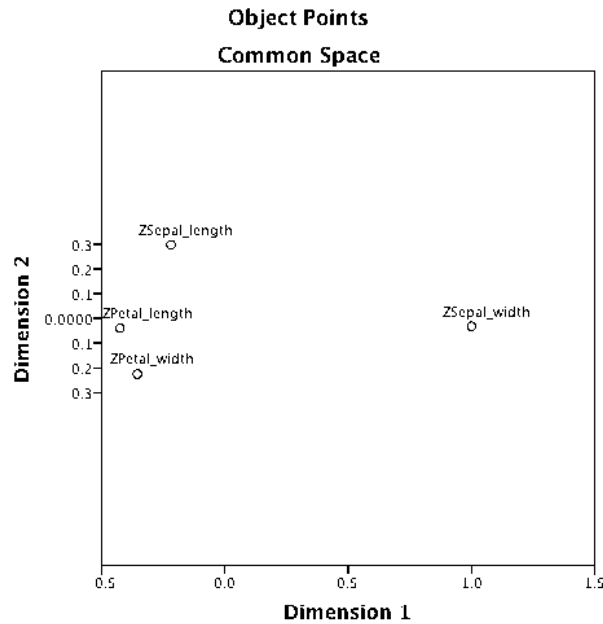


Fig. 4.1.2.2. The locations of the standardized variables according to the multidimensional scaling.

Table 4.1.2.1 also presents the communalities and eigenvalues, and fuzzy rule-based systems with seven rules and Takagi-Sugeno reasoning were used in an above-mentioned manner in this context. Our eigenvalues and communalities indicate that two components yield high loadings and all the variables are relevant in this context. These values also correspond to the PCA outcomes above. There is a slight negative correlation between C_1 and C_2 , and in this respect we have an oblique resolution.

Table 4.1.2.1. Intercorrelations between the standardized variables and principal components in the Iris data.

Variable	C_1	C_2	Communalities
ZSepal length	.999		1.000
ZSepal width		1.000	1.000
ZPetal length	.888		.944
ZPetal width	.838		.873
Eigenvalues	2.49	1.000	

Figure 4.1.2.3 depicts the locations of our variables in the principal component space and it also corresponds quite well to the PCA approach. Figure 4.1.2.4 depicts the locations of the variables as well as the principal components based on the PCA and our method when the multidimensional scaling is applied. We notice that our outcomes are slightly dissimilar to those of the PCA. In fact, our components are closer to singular variables and thus we should possibly fine-tune our model. This procedure is nevertheless precluded here because we have adopted the meta-level approach and our outcomes are already sufficiently plausible.

If we will generate the linear sum variable, $S_1 \approx C_1$, it would be similar to that of (4.1.1.1). The corresponding fuzzy-model approach, in turn, will base on the rule-based system, F_I ,

$$S_1 = F_I(SI, PI, Pw) \quad (4.1.2.1)$$

when C_1 is used as the dependent variable in the model construction.

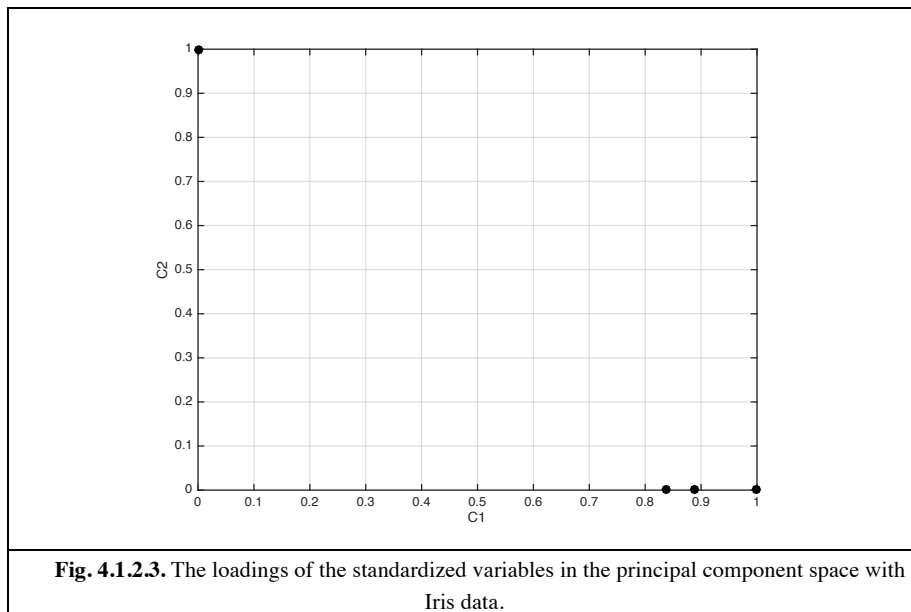
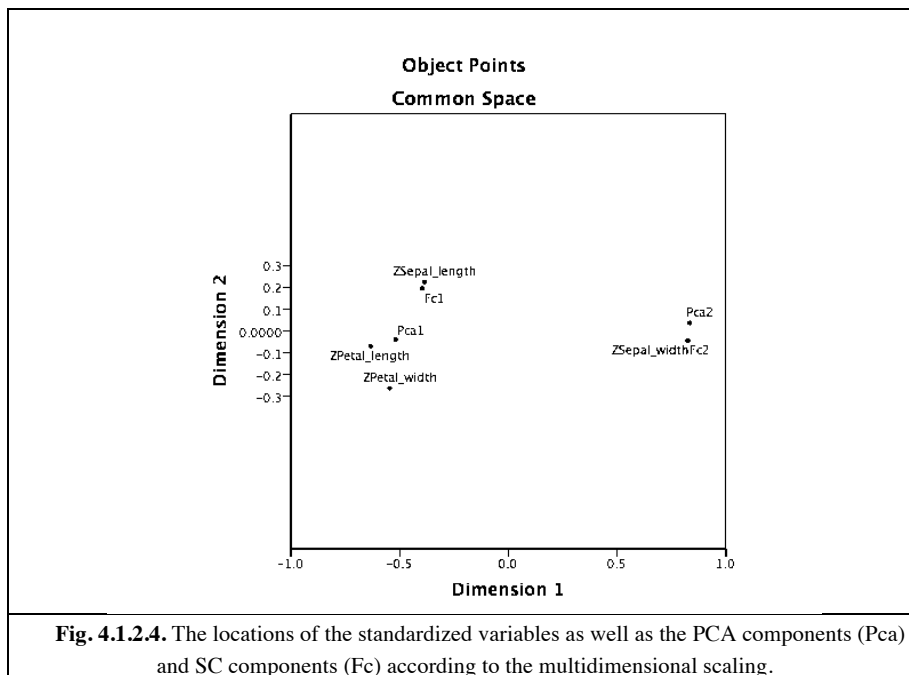


Table 4.1.2.2. Rmse values of the fuzzy models

$F_{ij}: C_i \rightarrow ZX_i$	
C1	C2
ZX1.042	.342
ZX2.282	.001
ZX3.099	.503
ZX4.134	.522



The intercorrelations between the PCA and our components are depicted in Fig. 4.1.2.5, and, as expected, they indicate high positive correlations.

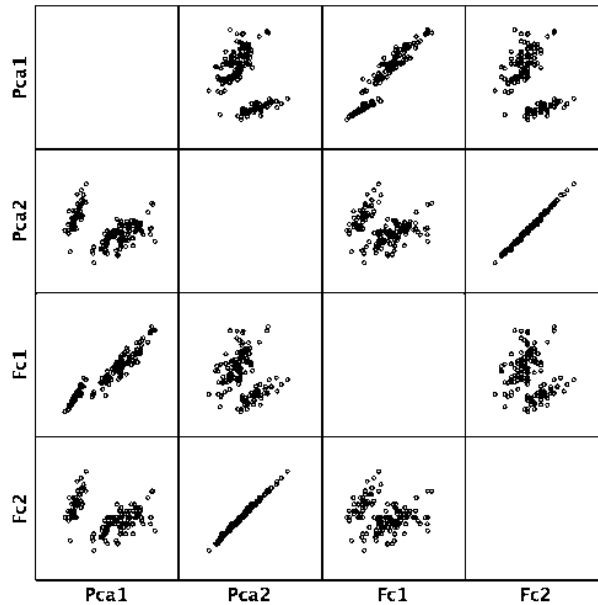


Fig. 4.1.2.5. Scatterplots of the PCA (Pca) and SC (Fc) components in the Iris data.

Since fuzzy systems are now applied, we could also establish that the closer the variables are to the components, the higher their degrees of membership, and vice versa. We could even replace our loadings with these memberships, if necessary. However, then the comparison between the distinct component extractions would be more difficult than in the case of correlations.

4.2. The World95 Data

Our second example deals with the benchmark data collected within the international world survey from 109 countries in 1995 (World95 data), and this is included in the SPSS example data sets, inter alia [34]. We focus on seven variables, *Average female life expectancy*, *Average male life expectancy*, *People who read (%)*, *Population increase (% per year)*, *Daily calorie intake*, *Log (base 10) of*

GDP per capita and *Birth to death ratio*. Fig. 4.2.1 depicts the inter-correlations between our variables.

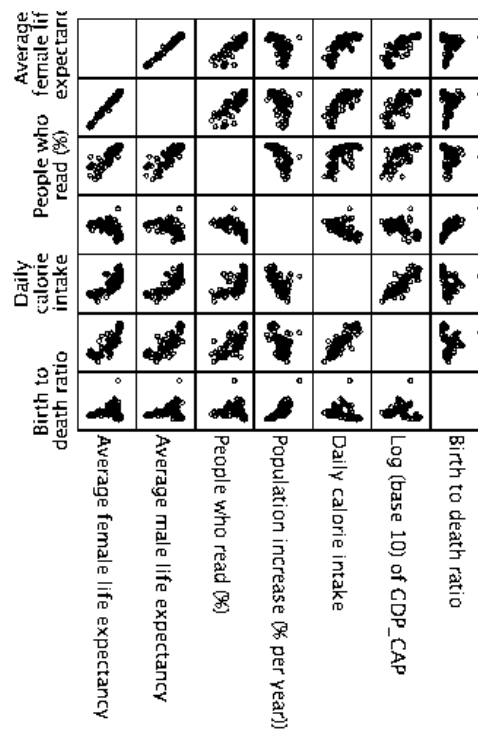


Fig. 4.2.1. Scatter plots of variables in the World95 data.

4.2.1. The PCA Approach

We used again PCA with Promax oblique rotation. Both the Kaiser-Meyer-Olkin measure and Bartlett's test now fulfilled the conditions on the satisfactory inter-correlations. The communalities were at least .91 and thus all the original variables seemed relevant in our analysis. We selected three components, even though only two components had eigenvalues higher than unity, because our decision seemed to reflect better the variable groups. These components explained approximately 95% of the total variance of the variables. The first two components have a quite high correlation (.753), and thus oblique rotation is justified.

Table 4.2.1.1 and Fig. 4.2.1.1 indicate that three variables have high loadings in the first principal component (the loadings less than .3 are omitted). The second and the third components seem to include two variables with high loadings. In our

outcome *Population increase* is not having a clear membership to any component. Since the foregoing table will not yield the correlations in oblique rotation, Table 4.2.1.2 presents the corresponding loading matrix based on the correlations between the variables and components. We notice that if we relied on this, the sum variable specification would be more problematic, but on the other hand, these values are better comparable to our SC-method outputs below.

Table 4.2.1.1. Rotated component matrix of Word95 data.

	Component		
	1	2	3
Average female life expectancy	.849		
Average male life expectancy	.830		
People who read (%)	1.000		
Population increase (% per year)	-.432		.759
Daily calorie intake		.951	
Log (base 10) of GDP/CAP		.767	
Birth to death ratio			1.000

Extraction Method: Principal Component Analysis.
 Rotation Method: Promax with Kaiser Normalization.

Table 4.2.1.2. The inter-correlations between the variables and the components in the World95 data.

	Component		
	1	2	3
Average female life expectancy	.973	.815	
Average male life expectancy	.955	.807	
People who read (%)	.955	.672	-.396
Population increase (% per year)	-.668	-.569	.892
Daily calorie intake	.733	.968	-.349
Log (base 10) of GDP/CAP	.800	.941	-.373
Birth to death ratio			.970

Extraction Method: Principal Component Analysis.
 Rotation Method: Promax with Kaiser Normalization.

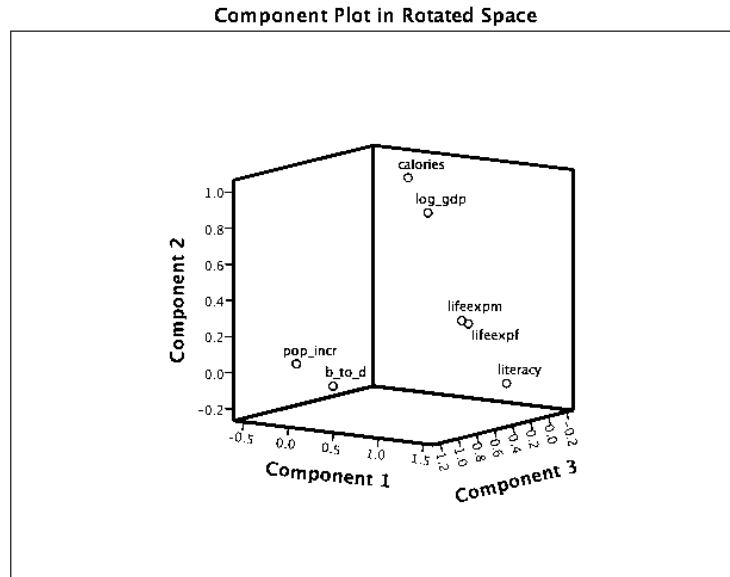


Fig. 4.2.1.1. Component plot in the rotated space with the World95 data.

If we specify now three sum variables, we may proceed as above by merely calculating the sums of those variables that have high loadings in the principal components. For example,

$$S_2 = \text{Daily calorie intake} + \text{Log (base 10) of GDP per CAP}, \quad (4.2.1.1)$$

or their averages, if the standard scores are used.

Hence, it seems plausible to specify sum variables among this data set even though now this task is more challenging than with the Iris data. Next we apply our SC method to this task.

4.2.2. The Soft Computing Approach

When our SC method is applied to three components, the correlation between the first two components is -0.565 , and thus in this respect we also apply an “oblique” method. According to multidimensional scaling, three cluster centers also seem plausible, even though clear clusters are now unavailable (Fig. 4.2.2.1).

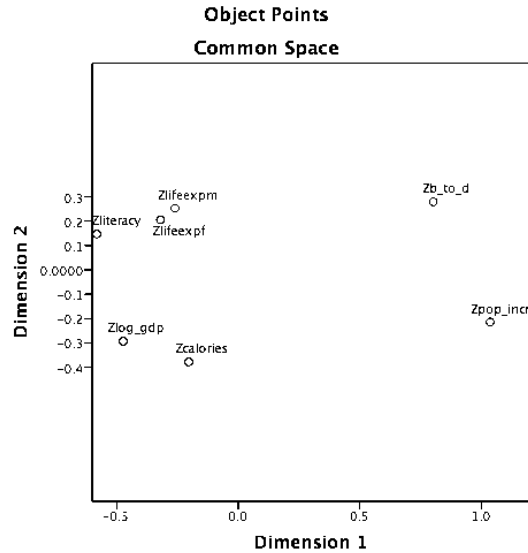


Fig. 4.2.2.1. The locations of the standardized variables according to the multidimensional scaling.

Our inter-correlations between the variables and the components seem somewhat distinct from the PCA outcomes with oblique loadings (Table 4.2.2.1, Figg. 4.2.2.2 and 4.2.2.3). Now the first component seems to comprise three variables common to both the PCA and SC approaches. In the SC model *ZLog (base 10) of GDP per CAP* is also having a high loading in the first component, and the same outcome is found in the correlation Table 4.2.1.2 above, and we must bear in mind that this table presents the correlations between the variables and the components in the manner of our SC method. The third component in PCA and the second component in the SC approach, in turn, provide quite similar outcomes.

Hence, the variables *ZLog (base 10) of GDP per CAP* and *ZDaily calorie intake* seem to yield distinct outcomes, but even in this case the correlation Table 4.2.1.2 corresponds quite well to our results. Fig. 4.2.2.4 depicts the scatter plots of our components. As regards our communalities (that were calculated according to the fuzzy models) and eigenvalues, they seem to fulfill the given conditions.

Table 4.2.2.1. Intercorrelations between the standardized variables and principal components in the World95 data.

Variable	C ₁	C ₂	C ₃	Communalities
ZAverage femal life expectancy	.999			.999
ZAverage male life expectancy	.982			.982
ZPeople who read (%)	.771			.795
ZPopulation increase (% per year)		.999		1.000
ZDaily calorie intake			1.000	1.000
ZLog (base 10) of GDP per CAP	.837			.876
ZBirth to death ratio		.816		.939
Eigenvalues	3.261	1.665	1.000	

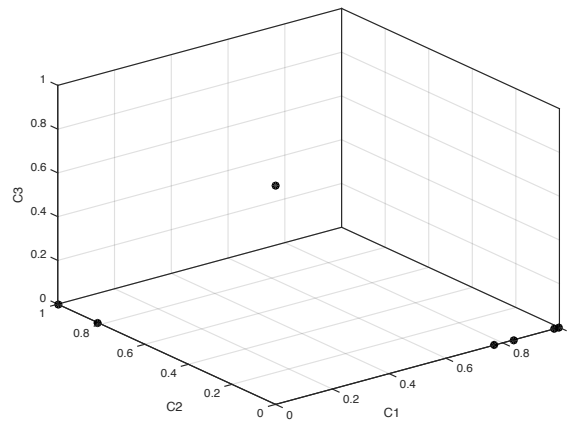


Fig. 4.2.2.2. The loadings of the standardized variables in the principal component space with World95 data.

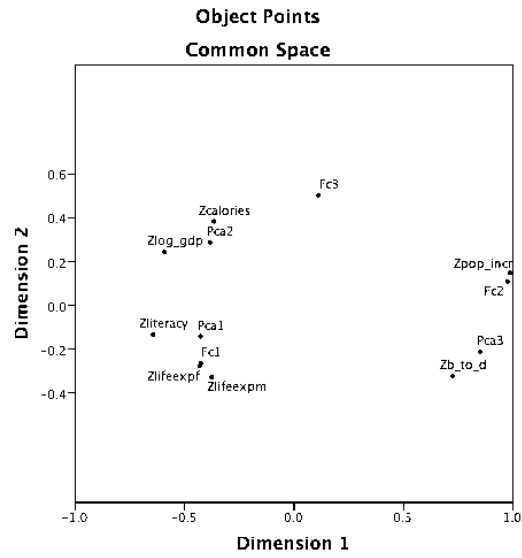


Fig. 4.2.2.3. The locations of the standardized variables as well as the PCA components (Pca) and SC components (Fc) according to the multidimensional scaling.

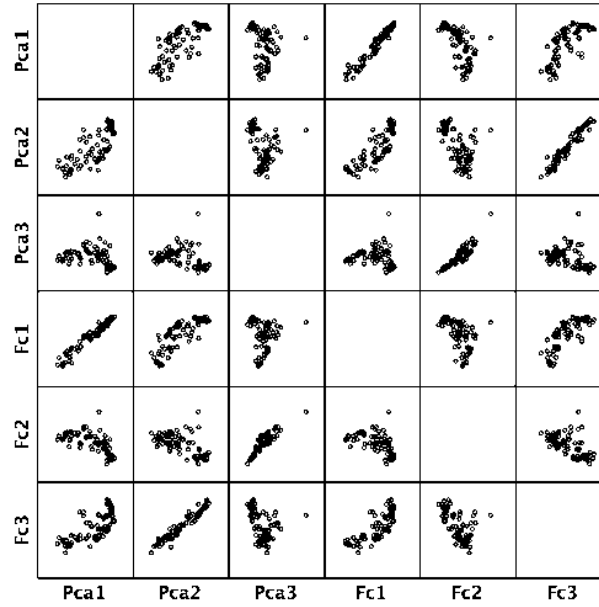


Fig. 4.2.2.4. Scatter plots of the PCA (Pca) and SC (Fc) components in World95 data.

Our sum variables, again, are the sums of the relevant variables in each component, for example,

$$S_2 = \text{Population increase (\% per year)} + \text{Birth to death ratio}, \quad (4.2.2.1)$$

or, by applying the corresponding fuzzy model,

$$S_2 = F_2(\text{Population increase (\% per year)}, \text{Birth to death ratio}), \quad (4.2.2.2)$$

with the dependent variable C_2 (Fig. 4.2.2.5).

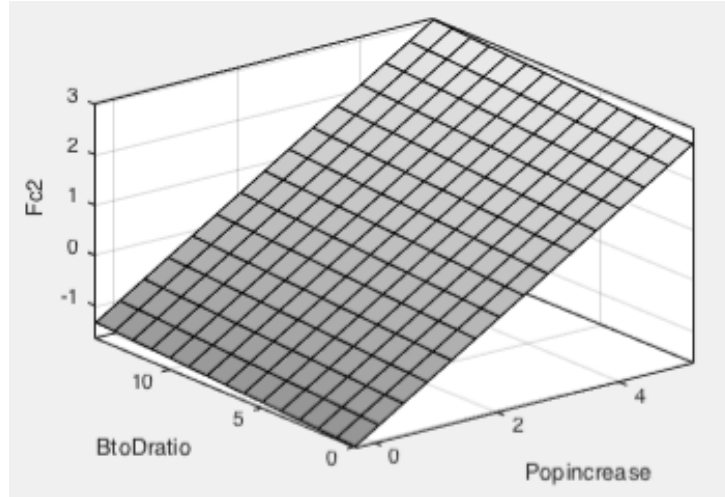


Fig. 4.2.2.5. A fuzzy model fitting to sum variable $S_2 = F(\text{Population increase (\% per year), Birth to death ratio})$.

Hence, in this context the SC method yields somewhat distinct outcomes, and this is due to our clustering approach and merely tentative calculations based on our general approach. In addition, the three-component approach seemed not fully justified in this context. On the other hand, in the human sciences the variables in the real world data often contain quite much noise and borderline cases.

5. Conclusions

We have considered above how the dimensionality reduction analogous to PCA may be carried out with fuzzy clustering method and fuzzy reasoning in an understandable manner, and two real-world data sets were also used as examples. Such prevailing traditional methods as PCA and FA are only appropriate to fairly limited usage because they presuppose linear correlations between the variables and normally distributed data sets, inter alia. Our SC approach, in turn, also seems usable to nonlinear and nonparametric data sets. The central idea in our approach is that we use fuzzy clustering method for finding the appropriate cluster centers to our variables, and these centers provide a basis for our sum variable construction.

In order to draw comparisons to the traditional approaches, our component loadings and goodness criteria based on various intercorrelations and *rsquares* between the variables and the principal components, and in this context we also ap-

plied fuzzy reasoning. However, in the long term we could replace the loadings with the degrees membership, as well as use even more fuzzy reasoning models. The loadings could also be linguistic values, if necessary. In this manner we could even better attain Lotfi Zadeh's recent idea on the fuzzy extended logic.

Since the SC community can already provide us with good model construction methods, we did not formulate any novel calculation technics but rather we focused at meta-level on constructing a tentative and an analogous system to PCA. Our contribution was to apply the fuzzy R-technique to data matrix and then construct fuzzy models for assessing the goodness of our outcomes with the loadings, communalities and eigenvalues. We also used fuzzy models for sum variable specifications because they are also appropriate to nonlinear cases.

We still have some open questions. First, due to the clustering approach, we encounter such their prevailing problems as the nonspherical clusters, selection of the correct metrics or the appropriate number of clusters. The lack of variable clusters or the great number of observations and original variables may also arouse problems.

Second, we still expect such standard goodness criteria for our dimensionality reduction within SC that may replace those of the PCA and FA. Examples of these are communalities and eigenvalues. Finally, if the degrees of membership are used for loadings, we still have various alternatives for specifying them.

One new frontier is to apply the fuzzy c-means clustering to this problem area, and this method would be analogous to FA. This is an interesting topic for the future studies and now it was mainly precluded due to the lack of space.

Despite the foregoing open problems, our SC approach seems nevertheless promising in practice as a "quick-and-dirty" method for the dimensionality reduction. However, further studies are still expected in this problem area.

References

- [1]. H. Bandemer, H. and W. Näther, W, "Fuzzy Data Analysis", Kluwer, Dordrecht, 1992.
- [2]. J. Bezdek, J. and S. Pal, "Fuzzy Models for Pattern Recognition", IEEE Press, New York, 1992.
- [3]. J. Bezdek & al., "Visual Assessment of Clustering Tendency for Rectangular Dissimilarity Matrices", IEEE Transactions on Fuzzy Systems, vol. 15/5, 2007, pp. 890-903.
- [4]. H. -M. Lee & al., "An Efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy", IEEE Transactions on Systems, Man and Cybernetics, Part B, Vol. 31/3, 2001, pp. 426-432.
- [5]. S. Chiu, "Fuzzy Model Identification Based on Cluster Estimation", Journal of Intelligent and Fuzzy Systems vol. 2, 1994, 267-278.
- [6]. A. Celikyilmaz, and B. Turksen, "Enhanced Fuzzy System Models with Improved Fuzzy Clustering Algorithm", IEEE Transactions on Fuzzy Systems 16/3, 2008, pp. 779-794.
- [7]. T. Denoeux and M. Masson, "Principal Component Analysis of Fuzzy Data Using Auto-associative Neural Networks", IEEE Transactions on Fuzzy Systems, vol. 12/3, pp. 336-349, 2004.

- [8]. H. Frigui and C. Hwang, "Fuzzy Clustering and Aggregation of Relational Data with Instance-Level Constraints", *IEEE Transactions on Fuzzy Systems*, vol. 16/6, 2008, pp. 1565-1581.
- [9]. P. Giordani and H. Giers, "A Comparison of Three Methods for Principal Component Analysis of Fuzzy Interval Data", *Computational Statistics & Data Analysis*, vol. 51, 2006, pp. 379-397.
- [10]. A. Gonzales and R. Perez, "Selection of Relevant Features in a Fuzzy Genetic Learning Algorithm", *IEEE Transactions on Systems, Man and Cybernetics, Part B*, Vol. 31/3, 2001, pp. 417-425.
- [11]. P. Grzegorzewski, & al. (Eds.), "Soft Methods in Probability, Statistics and Data Analysis", *Physica Verlag, Heidelberg*, 2002.
- [12]. J. Handl, and J. Knowles, "An Evolutionary Approach to Multiobjective Clustering", *IEEE Transactions on Evolutionary Computation*, vol. 11/1, 2007, pp. 56-76.
- [13]. H. Harman, "Modern Factor Analysis", *Univ. of Chicago Press, Chicago*, 1976.
- [14]. R. Hathaway and J. Bezdek, "Switching Regression Models and Fuzzy Clustering", *IEEE Transactions on Fuzzy Systems*, vol. 1/3, 1993, pp. 195-204.
- [15]. R. Hathaway and Y. Hu, "Density-Weighted Fuzzy C-Means Clustering", *IEEE Transactions on Fuzzy Systems*, vol. 17/1, 2009, pp. 243-253.
- [16]. T. -P. Hong and J. -P. Chen, "Finding Relevant Attributes and Membership Values", *Fuzzy Sets and Systems* 103/3, 1999, pp. 389-404.
- [17]. E. Hruschka, & al., "A Survey of Evolutionary Algorithms for Clustering", *IEEE Transactions on Systems, Man and Cybernetics, Part C*, 39/2, 2009, pp. 133-155.
- [18]. Y. Jin, "Fuzzy Modelling of High-Dimensional Systems: Complexity Reduction and Interpretability Improvement", *IEEE Transactions on Fuzzy Systems*, vol. 8/2, 2000, pp. 212-221.
- [19]. J. Kacprzyk and M. Fedrizzi, "Fuzzy regression analysis", *Physica Verlag, Heidelberg*, 1992.
- [20]. J. Kacprzyk, J. Owsinski, and D. Viattchenin, "A New Heuristic Possibilistic Clustering Algorithm for Feature Selection", *Journal of Automation, Mobile Robotics & Intelligent Systems* vol. 8, 2, 2014, pp. 40-46.
- [21]. F. Klawonn and J. Castro, J, "Similarity in Fuzzy Reasoning" *Mathware & Soft Computing*, 2/3, 1995, pp. 197-228.
- [22]. P. Luukka, "A New Nonlinear Fuzzy Robust PCA Algorithm and Similarity Classifier in Classification of Medical Data Sets", *Fuzzy Systems*, vol. 13/3, 2011, pp. 153-162.
- [23]. V. A. Niskanen, "Soft Computing Methods in Human Sciences", *Studies in Fuzziness and Soft Computing*, 134, Springer Verlag, Berlin, 2004.
- [24]. C. -H. Oh & al., "Fuzzy Clustering Algorithm Extracting Principal Components Independent of Subsidiary Variables", *Proceedings of the IEEE-INNS-ENNS*, vol. 3, 2000, pp. 377-380.
- [25]. N. R. Pal & al., "A Possibilistic Fuzzy C-Means Clustering Algorithm", *IEEE Transactions on Fuzzy Systems*, vol. 13/4, 2005, pp. 517-530.
- [26]. W. Pedrycz & al., "P-FCM: a Proximity-Based Fuzzy Clustering", *Fuzzy Sets and Systems* 146/1, 2004, pp. 21-42.
- [27]. M. Rezaee and A. Moini, "Reduction Method Based on Fuzzy Principal Component Analysis in Multi-Objective Possibilistic Programming", *The International Journal of Advanced Manufacturing Technology* vol. 67, 1-4, 2013, pp. 823-831.
- [28]. M. Smithson, "Fuzzy Set Analysis for Behavioral and Social Sciences", *Springer, New York*, 1986.
- [29]. P. Sneath, and R. Sokal, "Numerical Taxonomy", *Freeman, San Francisco*, 1973.
- [30]. T. Takagi and M. Sugeno, "Fuzzy Identification of Systems and Its Applications to Modeling and Control", *IEEE Transactions on Systems, Man and Cybernetics*, 15/1, 1986, pp. 116-132.

- [31]. Y. Yaeibuuchi and J. Watada, "Fuzzy Principal Component Analysis for Fuzzy Data", Proceedings of the Sixth IEEE International Conference, 1997, pp. 1127 – 1132.
- [32]. R. Yager, and D. Filev, "Generation of Fuzzy Rules by Mountain Clustering", Journal of Intelligent and Fuzzy Systems, vol. 2, 1994, pp. 209-219.
- [33]. T. Yang and S. Wang, "Robust algorithms for Principal Component Analysis", Pattern Recognition Letters, vol. 20, 1999, pp. 927-933.
- [34]. World95 data: <http://19-577-spring-2012.wiki.uml.edu/file/detail/World95.sav>
- [35]. M. -S. Yang & al., "Alpha-Cut Implemented Fuzzy Algorithms and Switching Regression", IEEE Transactions on Systems, Man and Cybernetics, Part B, vol. 38/3, 2008, pp. 588-603.
- [36]. L. Zadeh, "Fuzzy Logic = Computing with Words", IEEE Transactions on Fuzzy Systems, vol. 2, 1996, pp. 103-111.
- [37]. L. Zadeh, "Similarity Relations and Fuzzy Orderings", Information Sciences, vol. 3, 1971, pp. 177-200.
- [38]. L. Zadeh, "Toward Extended Fuzzy Logic – A First Step", Fuzzy Sets and Systems, vol. 160, 2009, pp. 3175-3181.
- [39]. J. Zar, "Biostatistical Analysis", Prentice-Hall, Englewood Cliffs, New Jersey, 1984.