

Clustering Variables Based on Fuzzy Equivalence Relations

Kingsley S. Adjenughwure, George N. Botzoris
and Basil K. Papadopoulos

Abstract We develop a method of grouping (clustering) variables based on fuzzy equivalence relations. We first compute the pairwise relationship (correlation) matrix between the variables and transform the matrix into a fuzzy compatibility relation. Then a fuzzy equivalence relation is constructed by computing the transitive closure of the compatibility relation. Finally, by taking all appropriate α -cuts, we obtain a hierarchical type of variable clustering. As examples, we use the proposed method first as a variable clustering tool in a regression model and secondly as a new way of performing factor analysis.

Keywords Fuzzy sets · Fuzzy equivalence relations · Clustering variables · Feature selection · Factor analysis · Regression analysis

1 Introduction

When modelling a system or a dependent variable, there is the problem of too many independent variables. In regression analysis for example, too many variables might lead to overfitting the data and render the model inefficient for the purpose of prediction. Also too many variables lead to a complex model and hence it is difficult to explain the interactions between all variables used in the model and how each variable affects the system (Sanche and Lonergan 2006). The problem of many variables is also known to affect cluster analysis. This is because for certain

K.S. Adjenughwure (✉) · G.N. Botzoris · B.K. Papadopoulos
Department of Civil Engineering, Democritus University of Thrace, 12, Vas. Sofias St.,
67100 Xanthi, Greece
e-mail: kingadje@civil.duth.gr

G.N. Botzoris
e-mail: gbotzori@civil.duth.gr

B.K. Papadopoulos
e-mail: papadob@civil.duth.gr

distributions, distances between data points become relatively uniform as the dimension increases (Beyer et al. 1999). Variable reduction is therefore useful when there are many observable variables and one need to choose a subset of these variables (feature selection) or find a reduced set of latent variables (dimensionality reduction) which explain a significant amount of the variance in the data.

Feature selection techniques are divided into wrapper, filters and embedded methods. Wrapper methods assign scores based on the predictive performance of each selected feature subset while filter methods use measures like correlation, mutual information, etc., to assign scores to feature subsets. Wrapper methods are known to be more accurate than filter methods but are computationally more expensive (Guyon and Elisseeff 2003). Embedded methods perform feature selection as part of the model calibration process. Their goal is to shrink the regression coefficients of all the variables in the model by constraining their sum to a specific value (Tibshirani 1996). This procedure effectively reduces most of the coefficients to zero leaving only the most important ones. Details about various variable selection techniques can be found in (Guyon and Elisseeff 2003). One obvious disadvantage of feature selection is that we only use a very few set of selected variables out of all the variables without analyzing the relationship between them.

Dimensionality reduction can be done directly by principal component analysis (PCA) or by first using K-means or other clustering techniques to partition the data into a fixed number of clusters and then projecting the data matrix into the space spanned by the centroid of the clusters (Karypis and Han 2000). Another method similar to PCA is factor analysis. This method is preferred when the goal is to describe the variance of a group of correlated variables with a single variable. This is equivalent to clustering groups of linearly related variables around a small number of latent variables called factors (Spearman 1904).

A disadvantage of dimensionality reduction techniques like PCA and factor analysis is that we cannot explicitly define the relationship between the latent (unobserved) variables and the observed variables. Also information is lost by using a few principal components instead of all components. There is also the problem of choosing how many latent variables or number of principal components. To solve the problems encountered by feature selection and dimensionality reduction, a method for variable reduction where all variables are retained is needed.

Variable clustering is a technique for reducing the number of variables to smaller set of clusters where the variables in a cluster are similar to other variables in the cluster and are dissimilar to other variables in another cluster. The similarity measure used here is that of correlation or other measures with the same meaning like mutual information. The advantage is that all variables are available for use after the clustering process. For example, the procedures of feature selection can be applied to the clustered variables by selecting variables from the clusters instead of individually (Bühlmann et al. 2013).

In this research, we introduce the idea of clustering variables using fuzzy equivalence relation. We also discuss the potential of the proposed model as both a feature selection and dimensionality reduction preprocessing tool. The contribution of this paper is to provide an alternative method to variable clustering and also give

new interpretations to the clustered variables and discuss their relationship with PCA, factor analysis, feature selection and high dimensional data clustering.

2 Literature Review

Although clustering variables is not as popular as data clustering, some advances have been made in this area. The first approach to variable clustering used principal component regression (Kendall 1957). Other variable clustering methods have been developed with regression analysis in mind. This process leads to an algorithm that performs variable clustering and model fitting at the same time. Examples of such algorithms are the one proposed by Dettling and Bühlmann (2004) and the OSCAR method developed by Bondell and Reich (2008).

Hastie et al. (2000) proposed the ‘gene shaving’ technique which uses principal components analysis to find a group of highly correlated variables. This method was applied to cluster genes and can either be unsupervised or supervised with a response variable. Hastie et al. (2001) also proposed the ‘tree harvesting’ method to select groups of predictive variables formed by hierarchical clustering in a supervised learning scheme.

For unsupervised variable clustering, other methods have been developed. PROC VARCLUS is an unsupervised variable clustering algorithm developed by SAS Institute Inc. This algorithm uses an iterative approach to find groups of variables that are as correlated as possible among themselves and as uncorrelated as possible with variables in other clusters. The algorithm begins with all variables in one cluster. A cluster is then chosen to split into two clusters by using the first two principal components and assigning the variables to the component with which they have the highest correlation (Nelson 2001).

Vigneau et al. (2001) used a method similar to PROC VARCLUS for variable clustering. In their approach, the goal was to maximize the correlation between the variables and their cluster centroid. The algorithm is similar to K-means and requires the user to specify the number of clusters. The authors also suggested a hierarchical variable clustering algorithm based on the same criteria to help with the selection of the number of clusters.

Palla et al. (2012) developed the Dirichlet Process Variable Clustering (DPVC). They clustered the variables by partitioning the observed dimensions by the Chinese Restaurant Process. DPVC exhibits the usual advantages over other methods because it is probabilistic and non-parametric. It can handle missing data, learn the appropriate number of clusters from data, and avoid overfitting.

Bühlmann et al. (2013) proposed a bottom-up agglomerative variable clustering algorithm based on canonical correlation. The algorithm starts with all variables in different clusters and successively merges the two clusters with the highest canonical correlation.

To our knowledge, fuzzy equivalence relation has not been applied to cluster variables. In this paper, we partition the variables into clusters using fuzzy equivalence

relations. Our approach is clearly linked to the data clustering application of fuzzy equivalence relations (Klir and Yuan 1995). All techniques used are the same as in data clustering. The only difference is that the distance or similarity measure used is the correlation among the variables. As noted earlier, we provide some new insight to the interpretation of the clustered variables by giving example of how the method can be used both as a standalone tool for unsupervised clustering of variables or as a preprocessing tool for variable selection in regression analysis, factor analysis, principal component analysis and clustering in high dimensional space. We consider that these new interpretations will be useful for other variable clustering methods too. The advantages of the proposed model over existing variable clustering methods are:

- Apart from using the correlation measures of the variables, the fuzzy proximity matrix in our proposed method can also be constructed using expert opinion on the relationship between the variables. This is very useful for variables that are not quantifiable or with missing, incomplete or unreliable data. The other methods mentioned above do not possess this flexibility, as they rely on the correlation calculations or other estimations made from the available data.
- There is no need to specify the number of clusters. The method produces a hierarchical cluster of all variables, if all α -cuts are used, and can also partition variables into an appropriate number of clusters, by choosing an appropriate α -cut most suitable for a particular application.
- The α -cut value used to cluster the variable represents the least correlation strength of all variables in the cluster. All variables therefore are equivalent at some level chosen by the user. This makes choosing one variable from the cluster or using the centroid of the variables as a cluster representative meaningful. Other existing methods do not have this property as they only find variables with the highest correlation to the centroid of the variables in the cluster or to other latent variable found by PCA.

3 Fuzzy Equivalence Relations

In the following we give some well-known notions and definitions.

Definition 1. Let X be a universal set. Every function of the form $A: X \rightarrow [0, 1]$ is called a fuzzy set or a fuzzy subset of X and $\mu_A(x)$ is the membership degree of x in the fuzzy set A .

Definition 2. The α -cut of the fuzzy set A is defined as the crisp set:

$${}^\alpha A = \{x \in R: \mu_A(x) \geq a\}, a \in (0, 1]$$

Definition 3. Let X, Y be universal sets. Then $R = \{[(x, y), \mu_R(x, y)] | (x, y) \in X \times Y\}$ is called a fuzzy relation on $X \times Y$.

Definition 4. Let $R \subset X \times Y$ and $S \subset Y \times Z$ be two fuzzy relations. The max-min composition $R \circ S$ is defined as:

$$R \circ S = \left\{ \left[(x, z), \max_y \left\{ \min_{x,z} \{ \mu_R(x, y), \mu_S(y, z) \} \right\} \right] \mid x \in X, y \in Y, z \in Z \right\}$$

Definition 5. A fuzzy relation R on $X \times X$ is called a fuzzy equivalence relation if it satisfies the following conditions:

1. Reflexive, that is $\mu_R(x, x) = 1, \forall x \in X$.
2. Symmetric, that is $\mu_R(x, y) = \mu_R(y, x), \forall x, y \in X$.
3. Transitive, that is $\mu_R(x, z) \geq \max_y \left\{ \min_{x,z} \{ \mu_R(x, y), \mu_R(y, z) \} \right\} \forall x, y, z \in X$.

Definition 6. A fuzzy relation on $X \times X$ is called a fuzzy compatibility relation if it is reflexive and symmetric.

Definition 7. The transitive closure R_T of the fuzzy relation R is the relation that is transitive, contains R and has the smallest possible membership grades.

Definition 8. Given a fuzzy compatibility relation R on $X \times X$, the transitive closure R_T can be calculated (Klir and Yuan 1995) by using the algorithm below:

1. $R' = R \cup (R \circ R)$.
2. If $R' \neq R$, make $R = R'$ and go to step 1.
3. Stop when $R' = R_T$.

The type of composition and set union in step 1 must be compatible with the definition of transitivity used. The max-min transitive closure corresponds to using the max-min composition and max operator for the set union.

Definition 9. The α -cut matrix of the fuzzy relation R is defined as:

$$R_\alpha = \left\{ \left((x, y), \mu_{R_\alpha}(x, y) \right) \mid \mu_{R_\alpha}(x, y) = 1, \text{ if } \mu_R(x, y) \geq \alpha, \right. \\ \left. \mu_{R_\alpha}(x, y) = 0, \text{ if } \mu_R(x, y) < \alpha, (x, y) \in X \times Y, \alpha \in [0, 1] \right\}$$

4 Proposed Method for Clustering Variables and Factor Analysis Based on Fuzzy Equivalence Relation

Suppose we have N variables to use to model a system, where N is considerably large. We form an $N \times N$ relation matrix R of the degree of ‘closeness’ between the variables and we express this degree of closeness in the interval $[0, 1]$. The proximity matrix of the variables is reflexive and symmetric and hence it is a fuzzy compatibility relation. The transitive closure R_T of the fuzzy compatibility relation is then

calculated using the algorithm in Definition 8. The new relation is now reflexive, symmetric and transitive and hence a fuzzy equivalence relation. The variables can then be clustered by choosing appropriate α -cuts. For an $M \times N$ data matrix X , where the rows represent data samples and the columns represent variables, the variable clustering procedure is summarized below.

Algorithm 1 – Variable clustering

Step 1 – Calculate the pairwise Pearson correlation (proximity) matrix of all variables.

Step 2 – Convert the matrix to fuzzy compatibility relation R by taking the absolute values of all entries. This is to make sure all values lie in the closed interval $[0, 1]$.

Step 3 – Find the transitive closure R_T of the fuzzy compatibility relation R using Definition 8.

Hierarchical clustering – Find all feasible clusters by using the α -cut matrices (Definition 9) of all unique values in R_T .

Partition clustering – Use prior knowledge about the variables or a cluster validity index to choose a suitable α -cut value that gives the best number of clusters.

Note that step 1 and step 2 above can be omitted if the fuzzy compatibility matrix is formed using expert knowledge about the variables. Also step 2 can be omitted if all values in R are in the interval $[0, 1]$. Finally any proximity metric which is reflexive and symmetric and whose values lie in the closed interval $[0, 1]$ can be used.

4.1 Empirical Example – Variable Clustering

We illustrate the variable clustering method using variables from a published regression analysis of the annual railway passenger demand of Greece (Profillidis and Botzoris 2005; Adjenughwure et al. 2013). There were a total of 12 independent variables in the analysis: average rail passenger travel distance, unit cost of rail transport, car ownership index, number of buses working in interurban routes, total bus vehicle-km travelled in interurban routes, average bus vehicle-km travelled in interurban routes, average bus passenger travel distance, unit cost of non-rail transport, the ratio of unit cost of bus transport to the unit cost of rail transport, cost of petrol, per capita Gross Domestic Product of Greece, and a variable which represents habitual inertia and constraints on supply.

The goal is to cluster the variables unsupervised and then use the results of the clustered variables for further analysis. To enhance the readers understanding of the method, we demonstrate a simple example with 7 independent variables.

Step1 –
Correlation matrix: $R = \begin{bmatrix} 1 & -0.68 & 0.73 & 0.77 & 0.74 & 0.58 & 0.60 \\ -0.68 & 1 & -0.68 & -0.85 & -0.76 & -0.57 & -0.55 \\ 0.73 & -0.68 & 1 & 0.72 & 0.57 & 0.38 & 0.95 \\ 0.77 & -0.85 & 0.72 & 1 & 0.77 & 0.51 & 0.59 \\ 0.74 & -0.76 & 0.57 & 0.77 & 1 & 0.94 & 0.43 \\ 0.58 & -0.57 & 0.38 & 0.51 & 0.94 & 1 & 0.26 \\ 0.60 & -0.55 & 0.95 & 0.59 & 0.43 & 0.26 & 1 \end{bmatrix}$

Step2 –
Absolute value of correlation matrix: $R = \begin{bmatrix} 1 & 0.68 & 0.73 & 0.77 & 0.74 & 0.58 & 0.60 \\ 0.68 & 1 & 0.68 & 0.85 & 0.76 & 0.57 & 0.55 \\ 0.73 & 0.68 & 1 & 0.72 & 0.57 & 0.38 & 0.95 \\ 0.77 & 0.85 & 0.72 & 1 & 0.77 & 0.51 & 0.59 \\ 0.74 & 0.76 & 0.57 & 0.77 & 1 & 0.94 & 0.43 \\ 0.58 & 0.57 & 0.38 & 0.51 & 0.94 & 1 & 0.26 \\ 0.60 & 0.55 & 0.95 & 0.59 & 0.43 & 0.26 & 1 \end{bmatrix}$

Step3 –
Transitive closure of R: $R_T = \begin{bmatrix} 1 & 0.77 & 0.73 & 0.77 & 0.77 & 0.77 & 0.73 \\ 0.77 & 1 & 0.73 & 0.85 & 0.77 & 0.77 & 0.73 \\ 0.73 & 0.73 & 1 & 0.73 & 0.73 & 0.73 & 0.95 \\ 0.77 & 0.85 & 0.73 & 1 & 0.77 & 0.77 & 0.73 \\ 0.77 & 0.77 & 0.73 & 0.77 & 1 & 0.94 & 0.73 \\ 0.77 & 0.77 & 0.73 & 0.77 & 0.94 & 1 & 0.73 \\ 0.73 & 0.73 & 0.95 & 0.73 & 0.73 & 0.73 & 1 \end{bmatrix}$

Example of partition clustering (α – cut matrix of R_T for $\alpha=0.8$): ${}^\alpha R_T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$

Each row or column corresponds to a variable. If any rows are equal, then they belong to the same cluster. From the matrix above, we find the clusters {1}, {2, 4}, {3, 7}, {5, 6}.

Example of hierarchical clustering: The unique value of R_T are 0.73, 0.77, 0.85, 0.94, 0.95. We form the α -cut matrix of R_T for each of these values and obtain the clusters below:

- {1, 2, 3, 4, 5, 6, 7} $\alpha=0.73$
- {1, 2, 4, 5, 6}, {3, 7} $\alpha=0.77$
- {1}, {2, 4}, {5, 6}, {3, 7} $\alpha=0.85$
- {1}, {2}, {4}, {5, 6}, {3, 7} $\alpha=0.94$
- {1}, {2}, {4}, {5}, {6}, {3, 7} $\alpha=0.95$
- {1}, {2}, {4}, {5}, {6}, {3}, {7} $\alpha=1.00$

The clusters produced for the 12 variables by the Algorithm 1 for $\alpha = 0.85$ is shown below:

$$\{1, 3, 7, 8, 11\}, \{2, 4\}, \{5, 6\}, \{9\}, \{10\}, \{12\}$$

As the researchers, now we have the clustered variables. We can select variables from the groups instead of individually as stated earlier. We can choose one or more variables from each group using the same feature selection criteria that we would use if the variables were not clustered. For instance, we can apply the variable selection technique to the first cluster. So instead of selecting among 12 variables $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, we reduce the problem to selecting among 5 variables $\{1, 3, 7, 8, 11\}$. This is then followed by selection among 2 variables $\{2, 4\}$ and $\{5, 6\}$.

This technique has two significant advantages: firstly, the computation cost of the feature selection algorithm is greatly reduced, and secondly, the grouping helps to reduce the problem of omitting some relevant variables since the set to choose from has been reduced. Note that the simplest way to select the variables is to pick one or more variable from each group which has the highest correlation with the dependent variable. This is of course not the best way as there are many other factors to be considered when selecting the variables. Luckily, there are many variable selection techniques any of which can be applied successively to the clustered variables to get the final set of variables. There are also available techniques specially made for selecting among clustered variables like the one used in PROC VARCLUS (Nelson 2001), or the one proposed by Bühlmann et al. (2013).

In our example of regression analysis with statistical criteria, the appropriate model was calibrated following the technique ‘general-to-specific model selection’ (Hendry 2000). The feature selection on all 12 variables with the dependent variables, gave the variables $\{2\}$, $\{3\}$, $\{9\}$, $\{10\}$, $\{11\}$, $\{12\}$ as the best calibration subset (Profillidis and Botzoris 2005; Adjenughwure et al. 2013). This is equivalent to choosing variables $\{3\}$ and $\{11\}$ from the first cluster, choosing variable $\{2\}$ from the second cluster, variables $\{9\}$, $\{10\}$ and $\{12\}$ from the single clusters and none from the cluster $\{5, 6\}$. We note that the criterion or method used for selecting the variables after clustering is left to the user. Our main aim for the paper is to propose an efficient method to quickly cluster variables unsupervised before the feature selection process begins.

Apart from the advantages previously listed, the proposed method is easier to implement compared to other variable clustering methods currently available. The only parameter to choose is α . A very small α will give fewer clusters with weakly correlated variables while a very large α will result in many clusters with highly correlated variables. If the goal is mainly to reduce the computational cost of feature selection algorithms, then use a small α to reduce the variables into few clusters and perform feature selection on each cluster. On the other hand, if we want to find or eliminate groups of highly correlated variables, we can use a bigger α . Note that the smallest α is the minimum value of the transitive closure matrix while the largest α is 1.

4.2 Empirical Example – Factor Analysis

We will use the proposed variable clustering based on fuzzy equivalence relation to perform factor analysis on the quality of service of public transport in Greece. The commuters’ perception on service quality offered by the public transport system of the city of Thessaloniki (Greece’s second-largest city) was recently measured by using customer satisfaction survey (nineteen closed-ended questions, five-point Likert scale answers, 450 respondents) and an exploratory factor analysis was performed to determine the principal components of service quality, (Delinasios 2014).

So, there are nineteen variables (the nineteen closed-ended questions) and our goal is to represent all these variables by using latent variables called factors. The procedure for factor analysis is almost the same as variable clustering with the only difference being that a single variable cannot form a cluster. Any cluster with at least two variables is called factor. The algorithm for factor analysis using fuzzy equivalent relation is described below.

Algorithm 2 – Factor analysis

Step 1 – Perform hierarchical variable clustering using the Algorithm 1 for variable clustering.

Step 2 (classical factor analysis) – Select the α -cut level with the highest number of factors and choose the smallest α -cut level with the same number factors as the selected.

Step 3 (variable assignment) – For each single non-assigned variable, use the correlation matrix to find the variable with which it has the highest correlation and assign the variable to the appropriate factor.

Step 4 (hierarchical factor analysis) – Repeat step 2 and step 3 for all unique number of factors to form a hierarchical type of factor analysis.

Step 1 – Hierarchical variable clustering:

{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19}	$\alpha=0.34$
{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 16, 17, 18}, {14, 15, 19}	$\alpha=0.40$
{1, 2, 3, 4, 16}, {5, 6, 7, 8, 9, 10, 11, 12, 13, 17, 18}, {14, 15, 19}	$\alpha=0.48$
{1, 2, 3, 16}, {4}, {5, 6, 7, 8, 9, 10, 11, 12, 13, 17, 18}, {14, 15, 19}	$\alpha=0.49$
{1, 2, 3, 16}, {4}, {5, 6, 7, 8, 9, 10, 17}, {11, 12, 13, 18}, {14, 15, 19}	$\alpha=0.56$
{1, 2, 3, 16}, {4}, {5, 7, 8, 9, 17}, {6}, {10}, {11, 12, 13, 18}, {14, 15, 19}	$\alpha=0.57$
.....	
{1}, {2}, {3}, {4}, {5}, {6}, {8}, {10}, {16}, {7, 9, 17}, {11, 12, 13, 18}, {14, 15, 19}	$\alpha=0.61$
.....	
{1}, {2}, {3}, {4}, {5}, {6}, {7}, {8}, {9}, {10}, ..., {15}, {16}, {17}, {18}, {19}	$\alpha=1.00$

Step 2 – Classical factor analysis: Remember that a factor is a group of two or more variables. From step 1, the highest number of factors is 4. There are many α -cut levels with four factors. We choose the smallest which is $\alpha = 0.56$.

$$\{1, 2, 3, 16\}, \{4\}, \{5, 6, 7, 8, 9, 10, 17\}, \{11, 12, 13, 18\}, \{14, 15, 19\} \quad \alpha = 0.56$$

Step 3 – Variable assignment: Only the 4th variable was separated. Using the correlation matrix, we find the variable to which {4} has the highest correlation. This is the {16}, so we put {4} in the same group as {16} and the final result is:

$$\{1, 2, 3, 4, 16\}, \{5, 6, 7, 8, 9, 10, 17\}, \{11, 12, 13, 18\}, \{14, 15, 19\}$$

This is exactly the factor analysis result with four factors as published by Delinasios (2014).

Step 4 – Hierarchical factor analysis: We only need to find the unique number of factors between the highest and the lowest number of factors from the α -cut levels. Since the highest number of factors is 4 and the lowest number of factors is 1. We repeat step 2 and step 3 of Algorithm 2 for number of factors 2 and 3. The smallest α -cut levels for 2 factors is $\alpha = 0.40$, while for 3 factors, $\alpha = 0.48$. The hierarchical factor analysis is presented below, (Fig. 1):

{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19}	1 factor,	$\alpha = 0.34$
{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 16, 17, 18}, {14, 15, 19}	2 factors,	$\alpha = 0.40$
{1, 2, 3, 4, 16}, {5, 6, 7, 8, 9, 10, 11, 12, 13, 17, 18}, {14, 15, 19}	3 factors,	$\alpha = 0.48$
{1, 2, 3, 4, 16}, {5, 6, 7, 8, 9, 10, 17}, {11, 12, 13, 18}, {14, 15, 19}	4 factors,	$\alpha = 0.56$

These are exactly the factor analysis results with 1, 2, 3, and 4 factors as produced by SPSS. Note that in step 2 of Algorithm 2, any number of factors can be selected as long as there is a corresponding α -cut level with the selected number of factors. As an advantage, *since* the α -cut levels are ordered, step 2 is equivalent to choosing the first α -cut level which has the same number of factors as the desired.

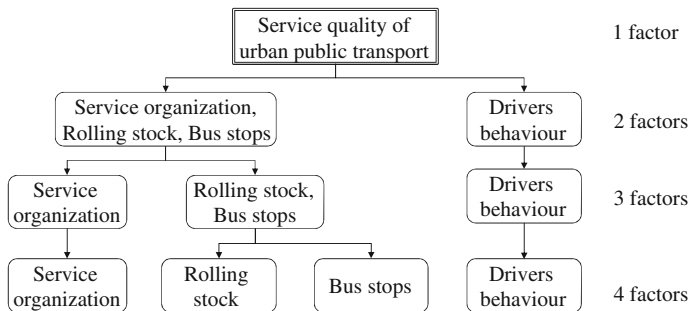


Fig. 1 The hierarchical factor analysis based on fuzzy equivalence relation

5 Some Interpretations of the Clustered Variables

Although the proposed algorithms are quite different from classical factor analysis and dimensionality reduction techniques, they share some similarities. Each cluster can be viewed as a principal component or a latent variable (factor) formed using the variables in that cluster. For example, we can form ‘factors’, ‘latent variables’ or ‘principal components’ by taking the average of the variables in each cluster (Vigneau et al. 2001). Therefore, the clustered variables can be used as a pre-processing tool to select the appropriate number of latent variables or principal components for the purpose of factor analysis or principal components analysis.

We note that PCA can also be performed separately on each cluster. To do this, we form the data matrix with rows as examples and columns as variables using only the variables in the cluster. We can then choose one or more components from the PCA of that cluster as the cluster representative. We underline that this last advantage of variable clustering has not been explored in details and it would be interesting to compare the components produced by the clustered variables with that of direct PCA using all variables.

The clustered variables can also be useful for clustering high dimensional data. This can be done either by doing data clustering using only variables in the same group (a subspace of the ‘closest’ variables) or using one variable from each variable cluster(subspace of the ‘farthest’ variables).

6 Conclusions

A variable clustering method based on fuzzy equivalence relation has been proposed. We have applied the method to cluster variables from a regression analysis. The results show that the method can be used to reduce the number of variables used for modeling a system and thus is an alternative to other variable clustering algorithm currently available. We have also modified the proposed method and used it for factor analysis. The results show that the method yields similar results to classical factor analysis.

The proposed method has some advantages over existing methods, as it does not require specification of number of clusters or of number of factors. Additionally, it can be implemented using expert opinion about the relationship of the variables. For regression purpose, the clustered variables are considered equivalent at some chosen level hence selecting one variable from each cluster or taking their average is justified. Finally, some interpretations of clustered variables have been offered to help the reader to better understand and use the results of variable clustering for other applications.

Acknowledgments This research was financially supported through the university scholarship for doctoral students (KE 81516) in the section of Mathematics and Informatics, Department of Civil Engineering. The scholarship is awarded by the Research Committee of the Democritus University of Thrace.

References

- Adjenughwure, K., Botzoris, G., Papadopoulos, B.: Neural, fuzzy and econometric techniques for the calibration of transport demand models. *J. Math. Sci. Eng. Appl.* **7**, 385–403 (2013)
- Beyer, K., Goldstein, J., Ramakrishnan, R., Shaft, U.: When is 'nearest neighbor' meaningful? In: 7th International Conference on Database Theory, pp. 217–235 (1999)
- Bondell, H.D., Reich, B.J.: Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR. *Biometrics* **64**, 115–123 (2008)
- Bühlmann, P., Rütimann, P., van de Geer, S., Zhang, C.-H.: Correlated variables in regression: clustering and sparse estimation. *J. Stat. Plann. Infer.* **143**, 1835–1858 (2013)
- Delinasios, N.: Parameterization of the Perceived by Users Quality of Service of Thessaloniki Urban Transport Organization. Democritus University of Thrace, Xanthi (2014)
- Detting, M., Bühlmann, P.: Finding predictive gene groups from microarray data. *J. Multivar. Anal.* **90**, 106–131 (2004)
- Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. *J. Mach. Learn. Res.* **3**, 1157–1182 (2003)
- Hastie, T., Tibshirani, R., Botstein, D., Brown, P.: Supervised harvesting of expression trees. *Genome Biol.* **2**, 1–12 (2001)
- Hastie, T., Tibshirani, R., Eisen, M., Alizadeh, A., Levy, R., Staudt, L., Chan, W.C., Botstein, D., Brown, P.: 'Gene shaving' as a method for identifying distinct sets of genes with similar expression patterns. *Genome Biol.* **1**, 1–21 (2000)
- Hendry, D.F.: The success of general-to-specific model selection. In: Hendry, D.F. (ed.) *Econometrics: Alchemy or Science? Essays in Econometric Methodology*, pp. 467–490. Oxford University Press, Oxford (2000)
- Karypis, G., Han, E.-H.: Fast supervised dimensionality reduction algorithm with applications to document categorization & retrieval. In: 9th Conference on Information and Knowledge Management, pp. 12–19 (2000)
- Kendall, M.: *A Course in Multivariate Analysis*. Griffin, London (1957)
- Klir, G.J., Yuan, B.: *Fuzzy Sets and Fuzzy Logic Theory and Application*. Prentice Hall PTR, New Jersey (1995)
- Nelson, B.D.: Variable reduction for modeling using PROC VARCLUS. In: 26th Annual SAS® Users Group International Conference, pp. 261–263 (2001)
- Palla, K., Knowles, D., Ghahramani, Z.: A nonparametric variable clustering model. In: 26th Annual Conference on Neural Information Processing Systems, pp. 2987–2996 (2012)
- Profillidis, V., Botzoris, G.: A comparative analysis of the forecasting ability of classical econometric and fuzzy models. *Fuzzy Econ. Rev.* **10**, 35–46 (2005)
- Sanche, R., Lonergan, K.: Variable reduction for predictive modeling with clustering. In: *Casualty Actuarial Society Winter Forum*, pp. 89–100 (2006)
- Spearman, C.: 'General intelligence', objectively determined and measured. *Am. J. Psychol.* **15**, 201–292 (1904)
- Tibshirani, R.: Regression shrinkage and selection via the Lasso. *J. Roy. Stat. Soc.* **58**, 267–288 (1996)
- Vigneau, E., Qannari, E.M., Punter, P.H., Knoops, S.: Segmentation of a panel of consumers using clustering of variables around latent directions of preference. *Food Qual. Prefer.* **12**, 359–363 (2001)