

## The Unfolding Analysis for Symbolic Objects Based on the Example of the External Car Advertisement Evaluation

**Artur Zaborski**

Wroclaw University of Economic and Business, Wroclaw, Poland

e-mail: artur.zaborski@ue.wroc.pl

ORCID: 0000-0003-1374-2268

**Marcin Pełka**

Wroclaw University of Economic and Business, Wroclaw, Poland

e-mail: marcin.pelka@ue.wroc.pl

ORCID: 0000-0002-2225-5229

© 2023 Artur Zaborski, Marcin Pełka

This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/4.0/>

**Quote as:** Zaborski, A., and Pełka, M. (2023). The Unfolding Analysis for Symbolic Objects Based on the Example of the External Car Advertisement Evaluation. *Econometrics. Ekonometria. Advances in Applied Data Analysis*, 27(4).

DOI: 10.15611/eada.2023.4.02

JEL Classification: C59, C81, M31

---

### Abstract

**Aim:** Multidimensional unfolding allows representing both columns (e.g. products, services) and rows (e.g. customers) of the preference matrix on the same low-dimensional map (usually it's a two or three-dimensional map). The main aim of the paper was to propose how to perform unfolding analysis for symbolic objects.

**Methodology:** The paper describes the possible ways of performing unfolding analysis for symbolic interval-valued data. The external unfolding is described in the details and used in the empirical part of the paper. The data (preferences and dissimilarities) were gathered by using the incomplete method of triads.

**Results:** The empirical part presents an application for unfolding symbolic data to evaluate customers' preferences, where car advertisements are used as the example. The results presented on a two-dimensional perceptual map allowed to discover seven groups of respondents with different preferences; most of them prefer Skoda, Audi, Volkswagen, and Honda advertisements to Toyota and Volvo.

**Implications and recommendations:** The proposed external approach for symbolic data allows to represent objects as rectangles (on two-dimensional map) or cuboids (in the case of three dimensions). The respondents are represented as points. Further work should focus on creating an algorithm that allows for the presentation of both symbolic objects and preferences expressed by respondents in the form of rectangles or cuboids.

**Originality/Value:** The paper presents an innovative and previously unrepresented external unfolding for symbolic data. Besides that it presents how other unfolding approaches could be adapted for symbolic data.

**Keywords:** symbolic data analysis, unfolding analysis, preference measurement, car advertisements

---

## 1. Introduction

Preferences are fundamental elements in the basic concept of the theory of economics and, in particular, in the consumer choice theory. They reflect consumers' attitudes developed in the process of mutual interactions between consumers and their environment. They are usually treated as a binary relationship based on axiomatic properties of reflexivity, transitivity, and consistency (e.g. Varian, 1997). Even though the relation of preferences is very easy to determine experimentally (e.g. using a questionnaire survey), the measurement aimed at quantifying preferences is problematic (Smoluk, 2000). There are no precise and unambiguous definitions of many concepts, therefore, it is difficult to measure both the intensity and level of the conditions described by these concepts. Modelling preferences, aimed at explaining the process of consumer behaviour, has been of interest to researchers since the 1960s. The modelling process involves estimating the structure of consumer preferences, i.e. assessing the usefulness for particular levels of attributes (variables), determining the relative significance of attributes, and analysing models that best reflect the description of multi-attribute objects (alternatives, selection options, profiles). The data used to estimate the structure of preferences generally originated from primary sources and relate to future decisions and choices to be made by consumers (this research of consumer preferences used primary data, stated preferences, and data from secondary sources – revealed preferences). Therefore, the modelling of the stated preferences followed three significant trends, which are compositional, decompositional, and mixed approaches.

The compositional approach uses the idea of the Fishbein-Rosenberg model of attitudes (Fishbein and Ajzen, 1975) primarily referring to the general utility for multi-attribute objects, in the additive model, as a weighted sum of the assessments of levels regarding the realisation of characteristics (attributes) in a given object (Green and Srinivasan, 1990; Zwerina, 1997). The basic method of preference analysis in the compositional approach is multidimensional preference scaling (including unfolding analysis) expressed in the form of graphic maps built in a reduced multidimensional space (Hair et al., 2006; Sagan, 2009).

In the decompositional approach, the respondents' preferences are used to determine the part-worth utilities for attribute levels, importance of attributes, etc. The decomposition approach uses mostly conjoint analysis methods (e.g. metric and nonmetric, additive models that can assume interactions which can be linear and nonlinear). The discrete choice models can also be used, which include binomial probability models (e.g. linear, logit, probit, complementary models) and polynomial (ordered and unordered categories) latent class models. The decomposition methods in preference studies are described in detail by Green and Srinivasan (1990), Bąk (2004; 2013).

The mixed (hybrid) approach is a combination of both techniques, i.e. compositional and decompositional. In this group, the models included combining the conjoint analysis model with models using data hierarchised directly by the respondents to estimate their preferences. Mixed

models have the adaptive version of conjoint analysis and hybrid models for conjoint analysis (Bağ, 2004, p. 44; Hensel-Börner and Sattler, 2000; Pieterse et al., 2010).

Preference measurement methods, included in particular groups, differ in the way of estimating the structure of preferences and the number of other factors, such as universality, simplicity, and the requirements to input data. However, the most frequently used criterion for assessing preference estimation models remains their predictive capacity.

Although the subject literature provides numerous studies on the predictive values of methods within the approaches mentioned previously (primarily various types of conjoint analysis), a still relatively small number of studies compare all the three types of preference estimation. There is no unambiguous evidence for the 'superiority' of any of these approaches in every research situation. Few studies discussing this matter indicate the comparable quality of decomposition and hybrid models and (in the case of some research) their slight superiority over compositional models (see e.g.: Green et al., 1981; Helm et al., 2004; Sattler and Hensel-Börner, 2007). Such an advantage (if it exists at all) is small enough to keep in mind that the choice between decomposition or hybrid and compositional models should also be determined by other factors than just their predictive capacity. Unfolding analysis, which is the basic method of preference analysis in the compositional approach, represents an alternative to the research on the structure of preferences in situations where time constraints or difficulties in obtaining relevant data from consumers reduce the other methods' value. The collection of data used in the unfolding analysis is not only less complicated, but also does not require such extensive assessment from the respondents as e.g. in the case of conjoint analysis, especially when the modelling process covers many attributes and their levels (see: Helm et al., 2004; Sattler and Hensel-Börner, 2007).

In a classical data situation, objects (patterns) are usually described by a vector of qualitative and quantitative measurements, where each column represents a single variable. However, the classical data situation is too restrictive to represent more complex data. To consider the uncertainty and/or variability of the data, the variables must assume a set of categories or intervals even with frequencies or weights. This type of data has been studied in Symbolic Data Analysis (SDA).

Symbolic data are of particular importance in preference studies. Collecting data in symbolic form enables the respondents to express their preferences in a more natural and fuller way, e.g. by providing a few preferred products or brands, and specifying several product features that affect their purchasing decisions. The indication by the respondents of only one preferred brand of the product, if there are more such brands, forced them to provide incomplete information. Collecting data in symbolic form makes measurement easier, because the respondents are more likely to indicate the range of expenses for the purchase of the product than a specific value. They may not know the exact value, prefer not to disclose certain information, or the value is difficult to estimate.

The main objective of SDA is to provide suitable methods and algorithms for dealing with aggregate or complex data where cells of the data contain sets of categories, intervals, or weights (distributions) (see: Billard and Diday, 2006; Bock and Diday, 2000; Diday and Noirhomme-Fraiture, 2008; Noirhomme-Fraiture and Brito, 2011).

The aim of this paper was to propose possible ways of performing unfolding analysis for symbolic interval-valued data and apply the external hybrid multidimensional unfolding to the data that reflects preferences toward car advertisements. The data (preferences and dissimilarities) were gathered by using the method of triads.

## 2. Principles of Unfolding Analysis

Unfolding for classical data attempts to produce configuration  $\mathbf{Y}$  of points in  $r$ -dimensional space with each point  $y_j$  ( $j = 1, \dots, m$ ) representing one of  $m$  judges, together with another configuration  $\mathbf{X}$  of

points  $x_i$  ( $i = 1, \dots, n$ ) in the same space, which represent choice objects. Individuals are represented as 'ideal' points in the multidimensional space so that the distances from each ideal point to the object points correspond to the preference scores. The ideal point model is used to find a point in a stimulus space, which is almost like an attribute. If the attribute is a subject's preference for the stimuli, then this point is interpreted as a subject's ideal stimulus. It is the hypothetical stimulus that the subject would prefer the most if it existed.

For preference judgements  $p_{ij}$  unfolding attempts to find configurations  $\mathbf{X}$  and  $\mathbf{Y}$  that minimise the STRESS function:

$$STRESS = \sqrt{\sum_{i<j} (d_{ij} - \hat{d}_{ij})^2 / \sum_{i<j} d_{ij}^2}, \quad (1)$$

where:  $d_{ij} = \sqrt{\sum_{a=1}^r (x_{ia} - y_{ja})^2}$  is the distance between  $x_i$  and  $y_j$ ,

$\hat{d}_{ij} = f(d_{ij})$  is a monotonic regression of  $d_{ij}$  on  $p_{ij}$ .

For non-metric preference judgements, disparities  $\hat{d}_{ij}$  must satisfy the monotonic restriction:

$$p_{ij} < p_{i'j'} \Rightarrow \hat{d}_{ij} \leq \hat{d}_{i'j'}.$$

There are two main approaches to unfolding procedure:

- internal unfolding,
- external unfolding.

In internal unfolding, both the object configuration and the ideal points are simultaneously derived only from the preference matrix. One can conceive the preference matrix as a submatrix of a dissimilarity matrix, in which the dissimilarity between the objects and between the respondents are treated as missing values (see Figure 1).

	objects	respondents
Objects	missing data	
Respondents		missing data

**Fig. 1.** Preference matrix as a submatrix dissimilarity matrix

Source: own work based on (Borg and Groenen, 2005).

The internal unfolding solution can be computed by the majorization algorithm, where STRESS is reduced by iteratively taking a Guttman transform. After  $K$  step of iteration, the updates of  $\mathbf{X}$  and  $\mathbf{Y}$  becomes (see Borg and Groenen, 2005):

$$\mathbf{X}^K = \mathbf{V}_{11}^+ [\mathbf{B}_{11}(\mathbf{X}^*, \mathbf{Y}^*) \mathbf{X}^* + \mathbf{B}_{12}(\mathbf{X}^*, \mathbf{Y}^*) \mathbf{Y}^*], \quad (2)$$

$$\mathbf{Y}^K = \mathbf{V}_{22}^+ [\mathbf{B}_{12}(\mathbf{X}^*, \mathbf{Y}^*)^T \mathbf{X}^* + \mathbf{B}_{22}(\mathbf{X}^*, \mathbf{Y}^*) \mathbf{Y}^*], \quad (3)$$

where:  $[V_{11}^+]_{n \times n} = m^{-1}(I - (n + m)^{-1}\mathbf{1}\mathbf{1}^T)$ ,

$[V_{22}^+]_{m \times m} = n^{-1}(I - (n + m)^{-1}\mathbf{1}\mathbf{1}^T)$ ,

$\mathbf{1}$  – column vector of ones,

$B_{12}(X^*, Y^*)$  – matrix with elements  $b_{ij} = \begin{cases} \frac{-\delta_{ij}}{d_{ij}(X^*, Y^*)} & \text{for } d_{ij}(X^*, Y^*) \neq 0 \\ 0 & \text{for } d_{ij}(X^*, Y^*) = 0 \end{cases}$ ,

$B_{11}(X^*, Y^*)$  – diagonal matrix with elements  $b_{ii} = -\sum_k b_{ik}$ ,

$B_{22}(X^*, Y^*)$  – diagonal matrix with elements  $b_{kk} = -\sum_i b_{ik}$ ,

$X^*, Y^*$  – configurations  $X$  and  $Y$  after  $K - 1$  step of iteration.

In external unfolding, it is assumed that a similarity objects configuration is given. With preference data on these objects, external unfolding puts the ideal point for each subject in the space, so that the closer this point lies to a point that represents an object, the more this object is preferred by an individual. The external analysis for the preference data is realised by PREFMAP (*PREFerence MAPping*), which consists of four preference-property models: vector model, simple unfolding model, weighted unfolding model and general unfolding model. Detailed algorithms for ideal points and vector models in PREFMAP are presented by Davison (1983).

### 3. Unfolding Analysis for Symbolic Data

In the case of symbolic objects, there are various ‘paths’ which can be used in an unfolding analysis which cover both an internal and an external unfolding analysis. Following the simplest classical approach, the symbolic data table can be transformed into classical data (e.g. by taking their midpoints, upper and lower bound of the interval, etc. (see e.g. de Souza et al., 2011; Diday and Noirhomme-Fraiture, 2008), and then unfolding analysis using the classical method can be performed (e.g. PREFMAP, PREFSCAL). Thus, the problem consists in the proper selection of methods for transforming symbolic variables into classical ones, and additionally this approach results in the loss of some information about symbolic objects.

Algorithms of unfolding analysis typically take distance matrices as input. From this point of view, their use for symbolic data described by variables of any type requires only calculating the appropriate distance for symbolic data (see Billard and Diday, 2006, pp. 231-248; Bock and Diday, 2000, pp. 166-183) and using them as classical data. Moreover, evaluation criteria (usually the STRESS coefficient) can be used for symbolic data, because they only analyse the relation between the distances in the original and reduced space. Hence, the second strategy is a variation of the classical strategy which consists in carrying out an analysis after prior application of the distance measure adequate for the symbolic data. In this case the loss of information about objects does not occur and the scaling results identify points representing symbolic objects. However, it seems problematic to present symbolic objects as points in the multidimensional space, due to the variables describing them, these objects do not take the form of points. The advantage of the presented methods is the ability to use the penalty function (see Busing et al., 2005) to avoid degenerate solutions.

The problem of presenting symbolic objects in the form of points, which occurs in the approaches presented earlier, is solved by the next two approaches. The first of them refers to external analysis and covers two stages. In the first stage, for the objects described using symbolic variables, the multivariate scaling is performed by applying one of the methods appropriate for this type of data (Interscal, I-Scal, SymScal).

In the Interscal method, which is an adaptation of the classical MDS, the initial interval-valued distances are replaced by a modified delta matrix that is defined as follows (Lechevallier, 2001, p. 54):

$$\tilde{\Delta} = \begin{bmatrix} 0 & 0 & \frac{\delta_{12}}{2} & \frac{\bar{\delta}_{12} + \delta_{12}}{2} & \dots & \frac{\delta_{1n}}{2} & \frac{\bar{\delta}_{1n} + \delta_{1n}}{2} \\ 0 & 0 & \frac{\bar{\delta}_{12} + \delta_{12}}{2} & \bar{\delta}_{12} & \dots & \frac{\bar{\delta}_{1n} + \delta_{1n}}{2} & \bar{\delta}_{1n} \\ \frac{\delta_{21}}{2} & \frac{\bar{\delta}_{21} + \delta_{21}}{2} & 0 & \bar{\delta}_{22} & \dots & \frac{\delta_{2n}}{2} & \frac{\bar{\delta}_{2n} + \delta_{2n}}{2} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ \frac{\delta_{n1}}{2} & \frac{\bar{\delta}_{n1} + \delta_{n1}}{2} & \frac{\delta_{n2}}{2} & \frac{\bar{\delta}_{n2} + \delta_{n2}}{2} & \dots & 0 & \frac{\delta_{nn}}{2} \\ \frac{\bar{\delta}_{n1} + \delta_{n1}}{2} & \bar{\delta}_{n1} & \frac{\bar{\delta}_{n2} + \delta_{n2}}{2} & \bar{\delta}_{n2} & \dots & \bar{\delta}_{nn} & 0 \end{bmatrix}, \quad (4)$$

where:  $\frac{\delta_{in}}{2}$  – the lower bound for  $i$ -th object and  $n$ -th variable,  
 $\bar{\delta}_{in}$  – the upper bound for  $i$ -th object and  $n$ -th variable,  
 $\frac{\bar{\delta}_{in} + \delta_{in}}{2}$  – the midpoint (centre) for  $i$ -th object and  $n$ -th variable.

The modified delta matrix is used to compute **B** matrix with elements (Lechevallier, 2001, p. 55):

$$b_{\alpha\beta} = -\frac{1}{2} \left( \tilde{\delta}_{\alpha\beta}^2 - \frac{1}{2n} \sum_{\gamma=1}^{2n} \tilde{\delta}_{\gamma\beta}^2 - \frac{1}{2n} \sum_{\eta=1}^{2n} \tilde{\delta}_{\eta\beta}^2 + \frac{1}{(2n)^2} \sum_{\gamma=1}^{2n} \sum_{\eta=1}^{2n} \tilde{\delta}_{\gamma\eta}^2 \right), \quad (5)$$

where:  $\alpha, \beta, \gamma, \eta = 1, \dots, 2n$  – number of elements of **B** and  $\Delta$  matrices.

Then, **B** matrix is used to obtain eigenvalues and eigenvectors. Eigenvectors and eigenvalues are used to get coordinates as follows (Lechevallier, 2001, p. 55):

$$x_{ua} = \sqrt{|\lambda_u|} v_{au}, \quad (6)$$

where:  $\lambda_1, \lambda_2, \dots, \lambda_{2n}$  – eigenvalues,  $v_1, v_2, \dots, v_{2n}$  – eigenvectors,  $u = 1, \dots, 2n$  – point in  $r$ -dimensional space,  $a = 1, \dots, r$  – number of dimensions.

Principal coordinates, which are interval-valued variables, are obtained as follows:

$$\underline{x}_{ia} = \min_{k \in \{2i-1, 2i\}} \{x_{ka}\}, \quad (7)$$

$$\bar{x}_{ia} = \max_{k \in \{2i-1, 2i\}} \{x_{ka}\}, \quad (8)$$

where:  $i = 1, \dots, n$  – number of symbolic object,  $a = 1, \dots, r$  – number of dimensions,  $x_{ka}$  –  $k$ -th coordinate from  $a$ -th dimension,  $\underline{x}_{ia}$  – lower bound of an interval-valued coordinate for  $i$ -th object and  $a$ -th dimension,  $\bar{x}_{ia}$  – upper bound of an interval-valued coordinate for  $i$ -th object and  $a$ -th dimension.

The SymScal (see Groenen et al., 2005) and I-Scal (see Groenen et al., 2006) methods use the idea of STRESS majorization for the interval-valued data. In the SymScal method, there is no normalised STRESS-Sym function, while in the case of I-Scal it is a normalised, within the range [0; 1] I-STRESS function. Critical steps regarding SymScal and I-Scal include selection number of iterations  $K$ , selection of matrix **X** for initial centres' coordinates for rectangles, selection of matrix **S** for ranges of rectangles, and selection of the stopping value  $\varepsilon$  (usually set to  $10^{-6}$ ).

The STRESS-Sym function used in the case of the SymScal method is expressed as follows (Groenen et al., 2006):

$$\text{STRESS-Sym}(\mathbf{X}, \mathbf{S}) = \sum_{i < k} \omega_{ik} [\underline{\delta}_{ik} - \underline{d}_{ik}(\mathbf{X}, \mathbf{S})]^2 + \sum_{i < k} \omega_{ik} [\bar{\delta}_{ik} - \bar{d}_{ik}(\mathbf{X}, \mathbf{S})]^2, \quad (9)$$

where:  $\mathbf{X}$  and  $\mathbf{S}$  – matrices of initial coordinates for centres ( $\mathbf{X}$ ) and ranges ( $\mathbf{S}$ ) of rectangles,  $\omega_{ik}$  – weights,  $\underline{\delta}_{ik}$  ( $\bar{\delta}_{ik}$ ) – minimal (maximal) dissimilarities between  $i$ -th and  $k$ -th symbolic object,  $\bar{d}_{ik}(\mathbf{X}, \mathbf{S}) = \sqrt{\sum_{a=1}^r [|x_{ia} - x_{ka}| + (s_{ia} + s_{ka})]^2}$  – maximal distance between symbolic objects,  $\underline{d}_{ik}(\mathbf{X}, \mathbf{S}) = \sqrt{\sum_{a=1}^r \max\{0, [|x_{ia} - x_{ka}| - (s_{ia} + s_{ka})]^2\}}$  – minimal distance between symbolic objects.

The normalised I-STRESS is expressed as follows (Groenen et al., 2006):

$$I - \text{STRESS}(\mathbf{X}, \mathbf{S}) = \frac{\sum_{i < k}^n \omega_{ik} [\underline{\delta}_{ik} - \underline{d}_{ik}(\mathbf{X}, \mathbf{S})]^2 + \sum_{i < k}^n \omega_{ik} [\bar{\delta}_{ik} - \bar{d}_{ik}(\mathbf{X}, \mathbf{S})]^2}{\sum_{i < k}^n \omega_{ik} \bar{\delta}_{ik}^2 + \sum_{i < k}^n \omega_{ik} \underline{\delta}_{ik}^2}, \quad (10)$$

where the elements of the normalised I-STRESS are similar to the notations from STRESS-Sym.

Detailed information on the majorization of STRESS-Sym and I-STRESS functions can be found in Groenen et al. (2006, pp. 18-26, 372-375), Groenen et al. (2005), and Dudek (2013, pp. 110-115).

The elements of  $\mathbf{X}$  and  $\mathbf{S}$  matrices can be selected in various ways. They can be the result of knowledge, expert opinions or can be obtained from the INTERSCAL method.

If objects are described with symbolic variables, they are presented in the form of intervals on a geometrically perceptual map. In two-dimensional space, the intervals are presented as rectangles and in the three-dimensional space as cubes.

In the second stage of the external unfolding analysis using preference data, such distribution of ideal points on the perceptual map is performed by means of the preference map method (e.g. PREFMAP) to match the distance of the ideal point from the symbolic objects with the respondents' preferences ordering, along with the ideal points distributed in relation to the centres of intervals.

The algorithm of external symbolic unfolding has the following stages:

1. Attainment of interval-valued variables for symbolic objects or collection of  $m$  judgments, opinions.
2. Performance of the multidimensional scaling for symbolic data with one of the methods: Interscal, SymScal, I-Scal.
3. Attainment of rectangles for the preference map.
4. Construction of matrix  $\mathbf{R}$  of rectangles' centres for unfolding analysis.
5. Collection of  $m$  preferences for  $n$  objects.
6. Mapping points representing the respondents through unfolding analysis for classical data (e.g. PREFMAP) on the configuration of rectangles centres (elements of matrix  $\mathbf{R}$ ).
7. Presentation of rectangles (representing columns) and points (representing rows) on one perceptual map.

The algorithm for internal unfolding was based on the idea of the SymScal and I-Scal algorithms for symbolic multidimensional scaling and the majorization of the proper I-STRESS function. The initial I-STRESS is defined by equation (10). For the classical data, one has to replace the initial dissimilarity matrix (see Figure 1) by matrix of objects  $\mathbf{X}$  and matrix of respondents  $\mathbf{Y}$  (for details on majorization of the STRESS function for classical unfolding see Groenen, 1993, p. 99). Then if one considers interval-valued preferences obtained either as intervals for individuals over time, intervals for more or less homogeneous groups of individuals or as intervals not for single products but groups of products, a similar unfolding analysis could be carried; in such cases both preferences and objects would be represented as rectangles.

Next,  $\underline{d}_{ik}(\mathbf{X}, \mathbf{S})$  and  $\bar{d}_{ik}(\mathbf{X}, \mathbf{S})$  (maximum distance between symbolic objects and minimum distance between symbolic objects) matrices should be replaced by two matrices for a lower bound distance and two for upper bound matrix, and the following majorization of the function can be performed in a similar way as for the classical STRESS-Sym function.

#### 4. Evaluation of Car Advertisements

To collect preference data for the presented study, the incomplete methods of triads were used (see Burton and Nerlove, 1976; Zaborski, 2017). The idea of the method is based on the theory of balanced incomplete block designs (see e.g. Cochran and Cox, 1957; Morris, 2010; Rink, 1987).

In the method of triads, the subject is asked to consider all possible groups of three objects ( $O_i, O_j, O_k$ ) ( $i, j, k = 1, 2, \dots, n$ , where  $i \neq j \neq k$ ) at a time, taken from the full set of  $n$  objects  $O = (O_1, O_2, \dots, O_n)$ . The subject has to indicate which two objects of each combination form the most similar pair and which two objects form the least matching pair. On this basis, the triad is created, where the most similar objects are placed as the first and the second, and the least similar to the first and the third one; for example, if  $(O_i, O_j)$  is the most similar pair and  $(O_j, O_k)$  the least similar pair, the triad is  $(O_i, O_j, O_k)$ .

An advantage of the triads method is the relative simplicity of the judgments required of the subjects. Although it is a useful technique for data collection, the number of triads increases very rapidly with the number of objects. When the number of triads is considered too large to be practical, according to the theory of balanced incomplete block designs, it can be reduced in such a way that all pairs of objects in it are presented equally frequently, but less often than their potential maximum number (see Burton and Nerlove, 1976; Roskam, 1970). Zaborski (2020) showed that the triad method gives satisfactory results if each pair of objects appears in the triad set at least twice, even when all the pairs of objects in triads cannot be presented equally frequently.

The creation of the triangular preference similarity matrix is possible by giving a pair of objects from the first and the second place in the triad two points, from the second and the third place one point, and from the first and the third place zero points. The value of an element  $p_{ij}$  in  $i$ -th row and  $j$ -th column of the matrix is the number of points awarded to a pair consisting of  $i$ -th and  $j$ -th objects in all triads.

To discover the perceptual map by using unfolding analysis, the similarity matrix should be transformed into a matrix of dissimilarities, especially if all pairs of objects in blocks cannot be presented equally frequently. Dissimilarities  $\delta_{ij}$  are determined in accordance with the formula:

$$\delta_{ij} = \begin{cases} 1 - \frac{p_{ij}}{2 \cdot m_{ij}} & \text{for } m_{ij} \neq 0 \\ 0,5 & \text{for } m_{ij} = 0 \end{cases} \quad (11)$$

where  $m_{ij}$  is the number of pairs  $(O_i, O_j)$  in the set of triads.

A total of 129 students from Lower Silesia in Poland (mainly from Jelenia Góra, Wrocław and Wałbrzych) evaluated six different (in terms of car brands) car advertisements (selected by the researchers) according to their preferences:

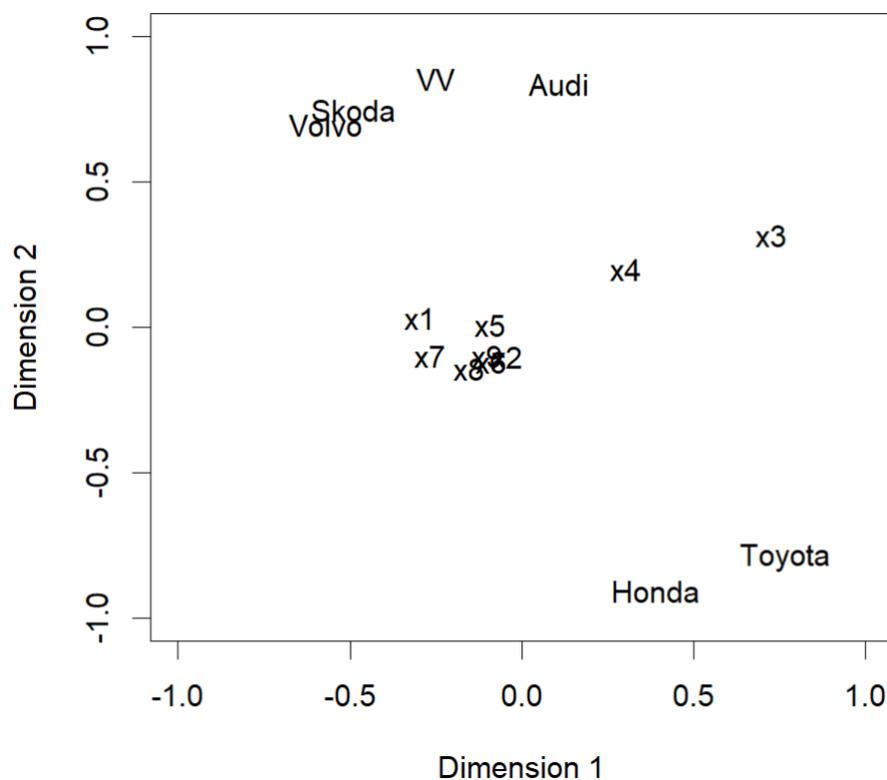
- Toyota Hybrid,
- Volvo XC90,
- Volkswagen id3,
- Honda e,
- Škoda Kodiaq,
- Audi e-tron.

Their task was to watch car advertisements on YouTube. The researchers provided them the links to all the advertisements, and then they filled in the questionnaire.

The respondents were asked to watch these advertisements and to rate each car advertisement according to nine features that could describe it on the seven-point scale, where 1 means "I totally agree" and 7 "I totally disagree", where:

- x1 – emotional,
- x2 – intelligent,
- x3 – entertaining,
- x4 – memorable,
- x5 – emphasising the lifestyle,
- x6 – trustworthy,
- x7 – targeting a certain group of customers,
- x8 – convincing,
- x9 – encouraging to buy.

The results obtained from the respondents were averaged and used in a traditional unfolding analysis to find out what elements describing car advertisements are located closer to some advertisements than others. The results are shown in Figure 2.



**Fig. 2.** Two-dimensional map representing car advertisements and the respondents' feelings concerning each of them

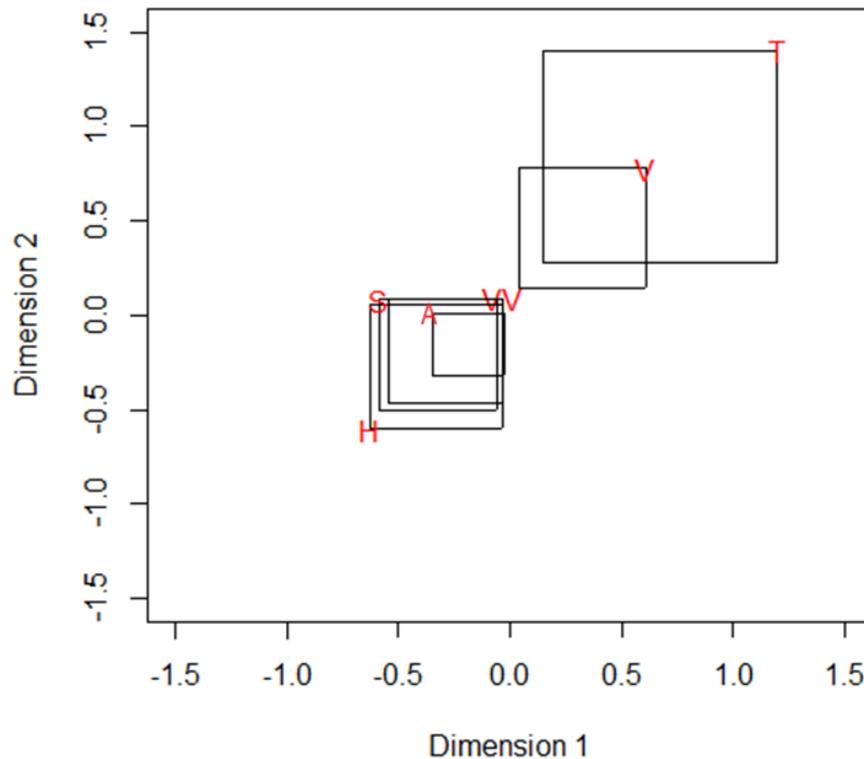
Source: own computation using R software.

The Audi car advertisement was considered to be memorable and entertaining, whereas the adverts for Volkswagen, Skoda and Volvo were seen as quite similar and located somewhat in similar distances from all the elements, but were not regarded as entertaining. The Toyota and Honda advertisements were also considered to be similar. When looking at the correlation between the variables and the car advertisements' characteristics, it can be said that they were regarded as more trustworthy, convincing, intelligent, less encouraging to buy and not targeting any specific group of customers.

For the purposes of the unfolding analysis, the students were asked (in line with the rules of the method of triads) to select the most similar and most dissimilar advertisements as well as the most and the least preferred advertisements. The gathered information was then used to obtain the dissimilarity matrix for each respondent and all advertisements, and also as the preference matrix for all respondents and advertisements. The distance matrices for each respondent were combined into

one final distance matrix, where the first quartile of the data set was used as the lower bound and the third quartile of the data set as the upper bound. As a result, symbolic interval-valued data distance was obtained. The SymScal algorithm was used to create a perceptual map for objects. As the initial (starting) values for matrices  $\mathbf{X}$  and  $\mathbf{S}$ , the results of Interscal method were used.

Figure 3 shows the perceptual map representing dissimilarities between the car advertisements, according to the evaluations obtained from the students.



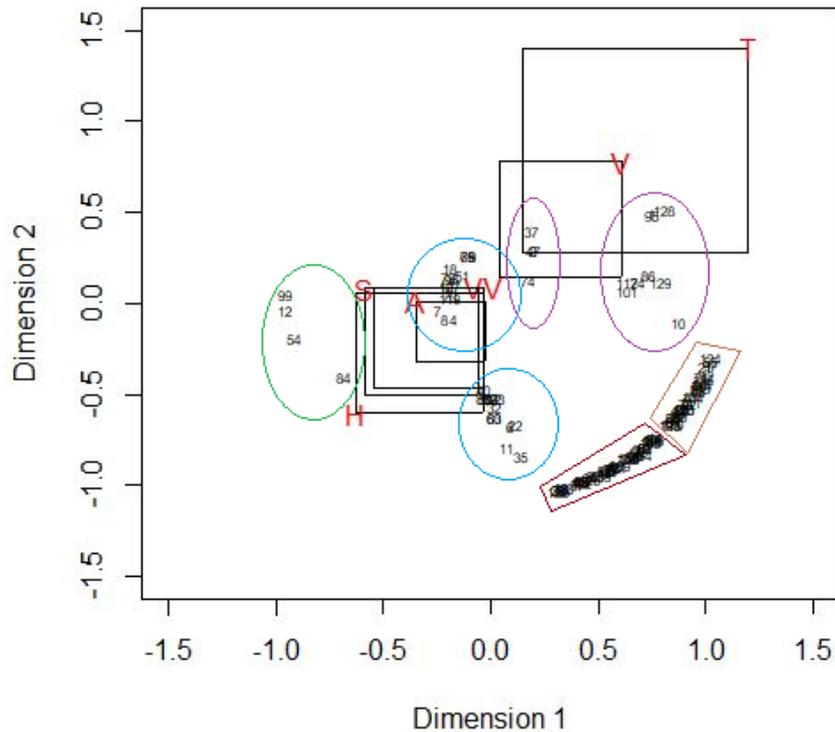
where: T – Toyota, V – Volvo, VV – Volkswagen, S – Skoda, A – Audi, H – Honda

**Fig. 3.** Two-dimensional map representing car advertisements

Source: own computation using R software.

Figure 3 also shows that two significant groups of objects can be distinguished. The first one includes the advertisements of Toyota and Volvo, which are similar to each other and quite dissimilar from the other car advertisements. The second group contains the advertisements of Skoda, Volkswagen, Honda and Audi which are similar according to the respondents. This may reflect the fact that Polish respondents pay much more attention to car brands, not engine type or other information that can be found in car advertisements. Polish customers believe that Toyota and Volvo are “the synonyms of trust and being reliable.”

Figure 4 presents the perceptual map representing both rows (the respondents) and columns (the car advertisements) of the preference matrix. The `symbolicDA` (Dudek et al., 2020) and `smacof` (Mair et al., 2020) R packages were used to obtain the results of symbolic multidimensional scaling and the results of the unfolding analysis.



where T – Toyota, V – Volvo, VV – Volkswagen, S – Skoda, A – Audi, H – Honda

**Fig. 4.** Two-dimensional map representing respondents and car advertisements

Source: own computation using R software.

Figure 4 shows that seven groups of the respondents can be distinguished. The green circle marks the first one, and the respondents from that group prefer Skoda, Audi, Volkswagen, and Honda advertisements to Toyota and Volvo. The blue circles mark the second two groups. The respondents from the higher blue circle group preferred the Audi advertisement to those of Volkswagen, Honda and Skoda. The students from this group showed a slight preference towards the Volvo car advertisement. The lower blue circle included the respondents with the same preferences towards Skoda, Audi, Honda, and Volkswagen car advertisements. They also liked less the advertisements by Toyota and Volvo. The violet circles mark the two next groups. The respondents from the first group preferred Volvo and Toyota car advertisements, but they also expressed a slight preference for the Audi e-Tron advertisement. The second group of respondents liked more advertisements by Volvo and Toyota than the others. Two polygons mark the two last groups. The upper one included the respondents with a slightly higher preference towards Volvo car advertisement, while the lower one includes respondents with similar preferences towards Honda, Audi, Skoda, and Volkswagen car advertisements who liked best advertisements by Volvo and Toyota.

## 5. Final Remarks

The article proposes how to conduct a multidimensional unfolding in the case of symbolic interval-valued data. In the hybrid approach, in the first step, the distances between symbolic objects were obtained, allowing to represent symbolic objects as rectangles. In the second step, preferences were added to the existing plot using centres of rectangles as starting, initial values. As a result, a low-dimensional (usually two or three-dimensional) map was created to analyse the preferences of the respondents.

This approach allows to collect all the information about the distances between the symbolic objects described by interval-valued variables, to take into account variability or uncertainty, and evaluate customers' preferences. However, this approach is limited and cannot be used directly to interval-valued preferences, yet such preferences could be transformed into classical data (to quantitative variables) with some information loss.

In terms of car advertisements, Audi advert was considered to be memorable and entertaining, whilst those for Skoda and Volvo were seen as quite similar and were placed in equal distances from all the car advertisements characteristics. Toyota and Honda were deemed to be more trustworthy, convincing, intelligent, less encouraging to buy, and not targeting any specific customer group.

When looking at the preferences and car advertisements, seven groups were identified. One can see that the respondents preferred advertisements for Skoda, Audi, Honda, Volkswagen cars to those for Volvo and Toyota.

The issue, which was only signalled in the article, requires further work by the authors to create a computer algorithm, and concerns the adaptation of I-STRESS for the purposes of the internal unfolding analysis for symbolic interval-valued preferences, with all the majorization steps that are needed to obtain final the rectangles/cuboids, both for objects and for 'ideal points'.

## References

- Bąk, A. (2004). *Dekompozycyjne metody pomiaru preferencji w badaniach marketingowych*. Wydawnictwo Akademii Ekonomicznej we Wrocławiu.
- Bąk, A. (2013). *Mikroekonometryczne metody badania preferencji konsumentów*. Wydawnictwo C.H. Beck.
- Billard, L., and Diday, E. (2006). *Symbolic Data Analysis. Conceptual Statistics and Data Mining*. John Wiley & Sons.
- Bock, H.-H., and Diday, E. (Eds.). (2000). *Analysis of Symbolic Data. Exploratory Methods for Extracting Statistical Information from Complex Data*. Springer-Verlag.
- Borg, I., and Groenen, P. (2005). *Modern Multidimensional Scaling. Theory and Applications*. Second Edition. Springer-Verlag.
- Burton, M. L., and Nerlove, S. B. (1976). Balanced Design for Triads Tests: Two Examples from English. *Social Science Research*, 5, 247-267. [https://doi.org/10.1016/0049-089X\(76\)90002-8](https://doi.org/10.1016/0049-089X(76)90002-8)
- Busing, F. M. T. A., Groenen, P. J. K., and Heiser, W. J. (2005). Avoiding Degeneracy in Multidimensional Unfolding by Penalizing on the Coefficient of Variation. *Psychometrika*, 1, 71-79. <https://doi.org/10.1007/s11336-001-0908-1>
- Cochran, W. G., and Cox, G. M. (1957). *Experimental Design*. (2nd ed.). Wiley.
- Davison, M. L. (1983). *Multidimensional Scaling*. John Wiley and Sons.
- De Souza, R. M. C. R., Queiroz, D. C. F., and Cysneiros, F. J. A. (2011). Logistic Regression-based Pattern Classifiers for Symbolic Interval Data. *Pattern Analysis and Applications*, 14(3), 273-282. <https://doi.org/10.1007/s10044-011-0222-1>
- Diday, E., and Noirhomme-Fraiture, M. (Eds.). (2008). *Symbolic Data Analysis with SODAS Software*. John Wiley & Sons.
- Dudek, A. (2013). *Metody analizy danych symbolicznych w badaniach ekonomicznych*. Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.
- Dudek, A., Walesiak, M., and Pełka, M. (2020). *Symbolic DA Package for R Software*. [www.r-project.org](http://www.r-project.org)
- Fishbein, M., and Ajzen, I. (1975). *Belief, Attitude, Intention and Behavior. An Introduction to Theory Research*. Addison-Wesley Publishing Company, Inc. Philippines.
- Green, P., Goldberg, S., and Montemayor, M. (1981). A Hybrid Utility Estimation Model for Conjoint Analysis. *Journal of Marketing*, 45, 33-41. <https://doi.org/10.2307/1251718>
- Green, P. E., and Srinivasan, V. (1990). Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice. *Journal of Marketing*, 54, 3-19. <https://doi.org/10.2307/1251756>
- Groenen, P. J. (1993). *The Majorization Approach to Multidimensional Scaling: Some Problems and Extensions*. DSWO Press. Leiden University.
- Groenen, P. J. F., Winsberg, S., Rodríguez, O., and Diday E. (2005). Multidimensional Scaling of Interval Dissimilarities. *Econometric Report, no. 2005-15*, Erasmus University.
- Groenen, P. J. F., Winsberg, S., Rodríguez, O., and Diday, E. (2006). I-Scal Multidimensional Scaling of Interval Dissimilarities. *Computational Statistics and Data Mining*, 51(1), 360-378. <https://doi.org/10.1016/j.csda.2006.04.003>

- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., and Tatham, R. L. (2006). *Multivariate Data Analysis* (Sixth edition), Prentice-Hall.
- Helm, R., Scholl, A., Manthey, L., and Steiner, M. (2004). Measuring Customer Preferences in New Product Development: Comparing Compositional and Decompositional Methods. *International Journal of Product Development*, 1, 12-29. <https://doi.org/10.1504/IJPD.2004.004888>
- Hensel-Börner, S., and Sattler, H. (2000). Validity of Customized and Adaptive Hybrid Conjoint Analysis. In R. Decker, et al. (Eds.), *Classification and Information Processing at the Turn of the Millennium* (pp. 320-329). Springer-Verlag.
- Lechevallier, Y. (Ed.) (2001). Scientific Report for Unsupervised Classification, Validation and Cluster Representation. *Analysis of Symbolic Official Data*. Project number IST-2000-25161, Technical Report for SODAS project.
- Mair, P., De Leeuw, J., Groenen, P. J. F., and Borg, I. (2020). *Smacof Package for R Software*. [www.r-project.org](http://www.r-project.org)
- Morris, M. (2010). Design of Experiments: *An Introduction Based on Linear Models*. Chapman and Hall/CRC.
- Noirhomme-Fraiture, M., and Brito, P. (2011). Far beyond the Classical Data Models: Symbolic Data Analysis. *Statistical Analysis and Data Mining*, 4(2), 157-170. <https://doi.org/10.1002/sam.10112>
- Pieterse, A. H., Berkens, F., Baas-Thijssen, M. C. M., Marijnen, C. A. M., and Stiggelbout, A. M. (2010). Adaptive Conjoint Analysis as Individual Preference Assessment Tool: Feasibility through the Internet and Reliability of Preferences. *Patient Education and Counseling*, 78(2), 224-233. <https://doi.org/10.1016/j.pec.2009.05.020>
- Rink, D. R. (1987). An Improved Preference Data Collection Method: Balanced Incomplete Block Designs. *Journal of the Academy of Marketing Science*, 15(3), 54-61. <https://doi.org/10.1177/009207038701500107>
- Roskam, E. E. (1970). The Methods of Triads for Multidimensional Scaling. *Nederlands Tijdschrift Voor de Psychologie*, 25, 404-417.
- Sagan, A. (2009). Analiza preferencji konsumentów z wykorzystaniem program Statistica – analiza conjoint i skalowanie wielowymiarowe. [https://media.statsoft.pl/\\_old\\_dnn/downloads/analiza\\_preferencji\\_konsumentow.pdf](https://media.statsoft.pl/_old_dnn/downloads/analiza_preferencji_konsumentow.pdf)
- Sattler, H., and Hensel-Börner, S. (2007). A Comparison of Conjoint Measurement with Self-explicated Approaches. In A. Gustafson, A. Herrmann (Eds.), *Conjoint Measurement*. [https://doi.org/10.1007/978-3-540-71404-0\\_4](https://doi.org/10.1007/978-3-540-71404-0_4)
- Smoluk, A. (Ed.) (2000). *Elementy metrologii ekonomicznej. Algebraiczne modelowanie pomiaru*. Wydawnictwo Akademii Ekonomicznej we Wrocławiu.
- Varian, H. R. (1997). *Mikroekonomia*. Wydawnictwo Naukowe PWN.
- Zaborski, A. (2017). The Influence of Triad Selection on the Preference Scaling Results. *Acta Universitatis Lodzianensis. Folia Oeconomica*, 4(330), 87-97. <https://doi.org/10.18778/0208-6018.330.06>
- Zaborski, A. (2020). The Use of the Incomplete Tetrad Method for Measuring the Similarities in Nonmetric Multidimensional Scaling. *Folia Oeconomica Stetinensia*, 20(1), 519-530. <https://doi.org/10.2478/fofi-2020-0030>
- Zwerina, K. (1997). *Discrete Choice Experiments in Marketing*. Physica-Verlag.

## **Analiza *unfolding* obiektów symbolicznych na przykładzie zewnętrznej oceny reklam samochodów**

---

### **Streszczenie**

**Cel:** Wielowymiarowa analiza *unfolding* pozwala na przedstawienie zarówno kolumn (np. produktów, usług), jak i wierszy (np. klientów) macierzy preferencji na tej samej mapie percepcyjnej (zwykle jest to mapa dwu- lub trójwymiarowa). Celem artykułu jest wskazanie propozycji przeprowadzenia analizy *unfolding* dla obiektów symbolicznych.

**Metodyka:** W artykule opisano możliwe sposoby przeprowadzenia analizy *unfolding* dla symbolicznych danych przedziałowych. Szczegółowo opisana zewnętrzna analiza *unfolding* została wykorzystana w części empirycznej artykułu. Dane (zarówno preferencje, jak i niepodobieństwa) zebrano z wykorzystaniem niepełnej metody triad.

**Wyniki:** W części empirycznej zaprezentowano możliwości zastosowania analizy *unfolding* dla danych symbolicznych w badaniu preferencji respondentów na przykładzie oceny wybranych reklam samochodów. Wyniki zilustrowane na dwuwymiarowej mapie percepcyjnej pozwoliły zidentyfikować siedem grup respondentów o różnych preferencjach względem przedstawionych reklam. Wyniki

badania wskazują, że dla większości respondentów reklamy Škody, Audi, Hondy i Volkswagena są bardziej preferowane niż reklamy proponowane przez Volvo i Toyotę.

**Implikacje i rekomendacje:** Zaprezentowane podejście do zewnętrznej analizy *unfolding* pozwala na prezentację obiektów w postaci prostokątów (w przestrzeni dwuwymiarowej) lub prostopadłościanów (w przestrzeni trójwymiarowej), a respondentów – w postaci punktów. Dalsze prace powinny skoncentrować się na stworzeniu algorytmu pozwalającego na prezentację zarówno obiektów symbolicznych, jak i wyrażanych przez respondentów preferencji w postaci prostokątów lub prostopadłościanów.

**Oryginalność/Wartość:** Artykuł prezentuje nowatorskie i nieprezentowane wcześniej podejście do zewnętrznej analizy *unfolding* dla danych symbolicznych. Ponadto przedstawia inne możliwe podejścia do symbolicznej analizy *unfolding*.

**Słowa kluczowe:** symboliczna analiza danych, analiza *unfolding*, pomiar preferencji, reklamy samochodów

---