
Mathematical Modelling and Analysis
Volume 23, Issue 4, 699–711, 2018
https://doi.org/10.3846/mma.2018.042

http://mma.vgtu.lt
ISSN: 1392-6292
eISSN: 1648-3510

On Mathematical Modelling of Synthetic Measures

Zbigniew Binderman[a], Bolesław Borkowski[a], Ryszard Kozera[a,b], Alexander N. Prokopenya[a] and Wiesław Szczesny[a]

[a] *Warsaw University of Life Sciences – SGGW*
Faculty of Applied Informatics and Mathematics
Nowoursynowska str. 159, 02-776 Warsaw, Poland
[b] *The University of Western Australia*
Computer Science and Software Engineering
35 Stirling Highway, WA 6009 Crawley, Perth, Australia
E-mail(*corresp.*): `boleslaw_borkowski@sggw.pl`
E-mail: `zbigniew_binderman@sggw.pl`

Received February 1, 2018; revised September 25, 2018; accepted September 26, 2018

Abstract. This work deals with some properties of synthetic measures designed to differentiate objects in a multidimensional analysis. The aggregate synthetic measures are discussed here to rank the objects including those validating the concentration spread. The paper shows that currently used various measures (based either on a single or a multiple model object) do not satisfy the necessary conditions requested to be met by a "good" synthetic measure.

Keywords: synthetic measure, development pattern, reverse problem.

AMS Subject Classification: 62P05; 91G70; 91G10.

1 Introduction

In economy applications Multidimensional Comparative Analysis (MCA) methods are commonly used to compare group of objects described by a set of n indicator variables (see [2, 3, 4, 5, 7]). The simplest application is ranking of the objects based on a certain non-directly observable variable. The latter is coined as either a synthetic measure or a development measure or a measure of concentration. Generally, in constructing such a variable for a given set of

economy objects the basic role is played by hypothetical model objects (see e.g. [1], [2], [11], [12], [13], [14], [15], [16], [18] or [20]). The additional task associated with object ordering is a high reduction of a large quantity of collected information about given objects into a few basic synthetically described categories (synthetic measures). Eventually these categories can be used as transformed input data for further analysis. It is worth noting that an inverse problem is potentially of significant importance, e.g. in medicine. Indeed the ordering or grouping the examined patients is well-known, where one searches for features and their correlated measures resulting in such ordering or grouping.

The main goal of any taxonomical analysis is grouping and ordering objects (units) of a multidimensional space. Various methods of classification and grouping are introduced to realize the above task (see [4,7,20]). The aim of this paper is to discuss basic properties of classification and grouping objects which should be satisfied by aggregate synthetic measures and to present possible difficulties in constructing such measures. Besides, we discuss here the new IC-concentration measures introduced in [7] and perform their analysis on well-posed problems [8], [9] or [10].

2 Selected techniques for creating synthetic measures

Ordering objects described by multiple features reduces into determination of relevant mapping $W : \mathbb{R}^k \to \mathbb{R}$ meeting the imposed constraints specified by a given recipient. Noticeably, it is not always possible for a recipient in question to define precisely such pertinent necessary conditions. A typical approach to construct the required transformation W (upon selecting its free variables) is to normalize all involved variables and potentially to replace them with the so-called stimulants. In doing so, the natural aim is to force all features/variables to be comparable (so that they all are positioned on the same scale). Having completed the above, one chooses next a suitable transformation W (called aggregated measure) which values are used to order a given set of input objects.

Relevant literature proposes a large number of indicators/measures - see e.g. [4], [5], [6] or [7]. Generally, construction methods are divided into two categories:

1. Those based on distance from a single or two models,
2. Those using utility functions or derived from intuition based on graphical representations of objects.

So far most of the proposed indicators in the literature are represented as functions of distance from one or two models. Note, that any synthetic measure W constructed by using distance from models can be normalized with the rescaled values ranging within the unit interval $[0, 1]$. The most essential items in the construction process of mapping W are introduced below.

Let $X = \mathbb{R}^n$ denote an n-dimensional vector space. Consider now a problem of ordering $m \in \mathbb{N}$ objects $\mathbf{Q}_1$, $\mathbf{Q}_2$,...,$\mathbf{Q}_m$ by using $n \in \mathbb{N}$ variables (features) meant to describe each of them. Without loss of generality, all features may be

considered as stimulants. Assume that symbol $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \in X$ (for $i = 1, 2, ..., m$) denotes a variables' vector describing the i-th object $\mathbf{Q}_i$. We say that $\mathbf{x}_i > \mathbf{x}_j$, (or $\mathbf{x}_i \geq \mathbf{x}_j$) (for $i, j = 1, ..., m$), if $x_{ik} > x_{jk}$ (or $x_{ik} \geq x_{jk}$), for $k = 1, 2, ..., n$. Besides, the notation $\mathbf{Q}_0$ and $\mathbf{Q}_{m+1}$ stands for the objects described by vectors with coordinates

$$x_{0,k} = \min_{1 \leq i \leq m} x_{ik}, \quad x_{m+1,k} = \max_{1 \leq i \leq m} x_{ik}, \quad k = 1, 2, ..., n, \tag{2.1}$$

respectively. Note that by (2.1) both vectors $\mathbf{x}_0$ and $\mathbf{x}_{m+1}$ can be treated as functions of $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m$, i.e.:

$$\mathbf{x}_0 = \lambda_1(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m) \quad \text{and} \quad \mathbf{x}_{m+1} = \lambda_2(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m). \tag{2.2}$$

Naturally, objects: $\mathbf{Q}_0$ described by vector $\mathbf{x}_0$ and $\mathbf{Q}_{m+1}$ described by vector $\mathbf{x}_{m+1}$ (perhaps fictitious) are not worse and not better, respectively, from the remaining objects $\mathbf{Q}_1$, $\mathbf{Q}_2$,...,$\mathbf{Q}_m$. Objects $\mathbf{Q}_0$ and $\mathbf{Q}_{m+1}$ can be treated as extra models added to the initial input objects $\mathbf{Q}_1, \mathbf{Q}_2, ..., \mathbf{Q}_m$. Upon simple inspection we have $\mathbf{x}_k \in \langle \mathbf{x}_0, \mathbf{x}_{m+1} \rangle$ ($\mathbf{x}_0 \leq \mathbf{x}_k \leq \mathbf{x}_{m+1}$ for $k = 1, 2, ..., m$).

A function d which maps a Cartesian product $X \times X$ into a set of non-negative numbers $\mathbb{R}^1_+ = \langle 0, +\infty)$ is said to represent *a distance* between any two elements $\mathbf{x}, \mathbf{y} \in X$, if it satisfies the following:

$$d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x}) \quad \text{and} \quad d(\mathbf{x}, \mathbf{x}) = 0.$$

The distance $d(\mathbf{x}, \mathbf{y})$ is called *a metric* if additionally d fulfills the triangle inequality:

$$d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}), \quad \text{for all} \quad \mathbf{x}, \mathbf{y}, \mathbf{z} \in X.$$

Given $\mathbf{x}, \mathbf{y} \in X$ with $\mathbf{x} = (x_1, x_2, ..., x_n)$ and $\mathbf{y} = (y_1, y_2, ..., y_n)$ the most widely known example of a distance (and a metric) reads as:

$$d_p(\mathbf{x}, \mathbf{y}) = \Big[\sum_{j=1}^{n} |x_j - y_j|^p \Big]^{1/p}, \quad 1 \leq p \leq \infty, \tag{2.3}$$

known as Minkowski's metric. On the other hand the following function:

$$d_{rad}(\mathbf{x}, \mathbf{y}) = \Big(\frac{1}{n} \sum_{i=1}^{n} |x_i - y_i||x_{i+1} - y_{i+1}| \Big)^{1/2} \tag{2.4}$$

with $x_{n+1} = x_1$ and $y_{n+1} = y_1$ satisfies merely a distance's but not a metric's axioms. To demonstrate the latter consider:

$$\mathbf{x} = (n, 1, 0, 0..., 0), \quad \mathbf{y} = (0, 0, ..., 0), \quad \mathbf{z} = (0, 1, 0, 0, ...0).$$

A simple inspection shows that:

$$d_{rad}(\mathbf{x}, \mathbf{y}) = 1, \quad d_{rad}(\mathbf{x}, \mathbf{z}) = 0, \quad d_{rad}(\mathbf{z}, \mathbf{y}) = 0.$$

Consequently

$$1 = d_{rad}(\mathbf{x}, \mathbf{y}) > d_{rad}(\mathbf{x}, \mathbf{z}) + d_{rad}(\mathbf{z}, \mathbf{y}) = 0.$$

Let the vectors $\mathbf{x}_0, \mathbf{x}_{m+1} \in \mathbb{R}^n_+$ be defined by formula (2.1) and satisfy the condition $\mathbf{x}_0 \neq \mathbf{x}_{m+1}$. Suppose that $\rho^*(\mathbf{x}, \mathbf{y})$ is a distance between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\rho^*(\mathbf{x}_0, \mathbf{x}_{m+1}) > 0$. The most commonly known synthetic indicators in the literature used to order objects are the following measures [7, 15]:

$$\begin{aligned}
\mu_1(\mathbf{x}) &= \frac{\rho^*(\mathbf{x}_0, \mathbf{x})}{\rho^*(\mathbf{x}_0, \mathbf{x}_{m+1})}, \quad \mu_2(\mathbf{x}) = 1 - \frac{\rho^*(\mathbf{x}_{m+1}, \mathbf{x})}{\rho^*(\mathbf{x}_0, \mathbf{x}_{m+1})}, \\
\mu_3(\mathbf{x}) &= \frac{\mu_1(\mathbf{x}) + \mu_2(\mathbf{x})}{2} = \frac{1}{2} + \frac{\rho^*(\mathbf{x}_0, \mathbf{x}) - \rho^*(\mathbf{x}_{m+1}, \mathbf{x})}{2\rho^*(\mathbf{x}_0, \mathbf{x}_{m+1})}, \\
\mu_4(\mathbf{x}) &= \frac{\mu_1(\mathbf{x})}{1 + \mu_1(\mathbf{x}) - \mu_2(\mathbf{x})} = \frac{\rho^*(\mathbf{x}, \mathbf{x}_0)}{\rho^*(\mathbf{x}_0, \mathbf{x}) + \rho^*(\mathbf{x}_{m+1}, \mathbf{x})},
\end{aligned} \tag{2.5}$$

where $\mathbf{x} \in \langle \mathbf{x}_0, \mathbf{x}_{m+1} \rangle$.

Visibly both measures μ_1 and μ_2 rely on a single model, whereas the remaining measures μ_3 and μ_4 depend on two models and are expressed as elementary functions of μ_1 and μ_2. Note also that by (2.2), all measures introduced in (2.5) are parameterized by m vectors:

$$\mu_i(\mathbf{x}) = \mu_1(\mathbf{x}; \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m), \tag{2.6}$$

where $i = 1, 2, 3, 4$.

For further consideration we normalize distances of vectors $\rho^*(\mathbf{x}, \mathbf{y})$, according to chosen model vectors $\mathbf{x}_0$, $\mathbf{x}_{m+1}$, by means of the following:

$$\rho(\mathbf{x}, \mathbf{y}) = \rho^*(\mathbf{x}, \mathbf{y}) / \rho^*(\mathbf{x}_0, \mathbf{x}_{m+1}).$$

Upon renormalization $\rho(\mathbf{x}_0, \mathbf{x}_{m+1}) = 1$. In addition, (2.5) reformulates into:

$$\begin{aligned}
\mu_1(\mathbf{x}) &= \rho(\mathbf{x}_0, \mathbf{x}), \quad \mu_2(\mathbf{x}) = 1 - \rho(\mathbf{x}_{m+1}, \mathbf{x}), \\
\mu_3(\mathbf{x}) &= \frac{1}{2}\left[1 + \rho(\mathbf{x}_0, \mathbf{x}) - \rho(\mathbf{x}_{m+1}, \mathbf{x})\right], \\
\mu_4(\mathbf{x}) &= \rho(\mathbf{x}_0, \mathbf{x}) / \big(\rho(\mathbf{x}_0, \mathbf{x}) + \rho(\mathbf{x}_{m+1}, \mathbf{x})\big).
\end{aligned} \tag{2.7}$$

In the special case when models $\mathbf{x}_0 = \mathbf{0} = (0, 0, ..., 0)$, $\mathbf{x}_{m+1} = \mathbf{1} = (1, 1, ..., 1)$ then

$$\rho^*(\mathbf{0}, \mathbf{1}) = \begin{cases} 1, & \text{for } \rho^* = d_{rad}, \\ n^{1/p}, & \text{for } \rho^* = d_p. \end{cases}$$

It is easy to see that the considered measures, defined by (2.7) are *normalized* in relationship to chosen models, i.e.:

$$\mu_i(\mathbf{x}_0) = 0 \quad \text{and} \quad \mu_i(\mathbf{x}_{m+1}) = 1, \quad \text{for} \quad i = 1, 2, 3, 4.$$

Let vector $\mathbf{x} \in \mathbb{R}^n_+$ be now arbitrarily fixed. With the aid of μ_1 and μ_2 can construct alternative measures, possibly according to (see, e.g., [14]):

$$\begin{aligned}
\mu_5(\mathbf{x}) &= \left\{ \begin{array}{ll} \frac{2\mu_1(\mathbf{x})\mu_2(\mathbf{x})}{\mu_1(\mathbf{x})+\mu_2(\mathbf{x})}, & \text{for } \mu_1(\mathbf{x})+\mu_2(\mathbf{x}) \neq 0, \\ 0, & \text{for } \mu_1(\mathbf{x})+\mu_2(\mathbf{x}) = 0. \end{array} \right\} - \text{harmonic mean}, \\
\mu_6(\mathbf{x}) &= \sqrt{\mu_1(\mathbf{x})\mu_2(\mathbf{x})} - \text{geometric mean}, \\
\mu_7(\mathbf{x}) &= \sqrt{0.5\big(\mu_1^2(\mathbf{x})+\mu_2^2(\mathbf{x})\big)} - \text{root mean square}.
\end{aligned} \tag{2.8}$$

Again by (2.2) all measures from (2.8) are parameterized by m vectors:

$$\mu_i(\mathbf{x}) = \mu_1(\mathbf{x}; \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m),$$

for $i = 5, 6, 7$. It can be shown (see [14]) that for each $\mathbf{x} \in \langle \mathbf{x}_0, \mathbf{x}_{m+1} \rangle$ the following inequalities hold:

$$\min\{\mu_1, \mu_2\} \leq \mu_5 \leq \mu_6 \leq \mu_3 \leq \mu_7 \leq \max\{\mu_1, \mu_2\}.$$

It is easily visible that the most common synthetic indicator w_i defined as an arithmetic mean of normalized values of variables can be expressed as a simple dependence on the distance from the negative model $\mathbf{x}_0$. In the case of normalization known as zero unitarization (see [4]) this synthetic indicator reads:

$$w_i = \frac{1}{n}\sum_{j=1}^{n} z_{ij} = \frac{\rho^*\big((0, ..., 0), (z_{i1}, ..., z_{in})\big)}{\rho^*\big((0, ..., 0), (1, ..., 1)\big)},$$

or in the case of normalization done by standardization it coincides with:

$$w_i = \frac{1}{n}\sum_{j=1}^{n} z_{ij} = \frac{1}{n}\left(\rho^*\big((z_{01}, ..., z_{0n}), (z_{i1}, ..., z_{in})\big) - \rho^*\big((0, ..., 0), (z_{01}, ..., z_{0n})\big)\right),$$

where $\mathbf{z}_i$ denote vectors $\mathbf{x}_i$ after normalization, $i = 0, 1, ..., m+1$.

Having a vast, practically limitless, set of functions that can be used as distances, analysts can create rich sets of measures useful in ordering of objects described by variables in automated reporting systems. Thus an important issue unifying such indicators is determination of measures used.

3 Properties of basic synthetic measures

For cyclical reporting (e.g. in automated reporting systems) ordering objects is an important issue. The report creator concentrates here on both universal properties of the synthetic indicator that orders input objects and on the specific properties of the concrete situation. Universal properties include, among others, invariability in relation to changes in scale when all variables have values on an interval scale (or only on a quotient scale) or in relation to the indicator assuming a value a when the vector that describes the object, has after normalization, only values of a as coordinates. Indeed, if μ denotes a synthetic

measure of an object which is determined by the vector $\mathbf{a}$ then $\mu(\mathbf{a}) = a$, where $a \geq 0$.

On the other hand, an example of a specific property is a requirement to have an increase in the value of a given indicator depending on component variable, after normalization, has its value change by a. Moreover, it is often useful to the user to know in which situations values of indicators overlap and in which they significantly differ. Properties of commonly used indicators that are presented bellow partially answer the above questions.

Property 1. Assume the indicators considered are treated as functions of $\mathbf{x} \in \langle \mathbf{x}_0, \mathbf{x}_{m+1} \rangle$ and vectors defining objects $\mathbf{Q}_1$, $\mathbf{Q}_2$,...,$\mathbf{Q}_m$ (where for $i = 1, 2, 3, 4$ define $\mu_i = \mu_i(\mathbf{x}; \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m)$). Take the distance $\rho^* = d_p$ as introduced in (2.5). Then each measure μ_i $(i = 1, 2, 3, 4)$ is a homogeneous function of the zero order.

Proof. Let $\alpha > 0$ and $\mathbf{x} = (x_1, x_2, ..., x_n) \in \langle \mathbf{x}_0, \mathbf{x}_{m+1} \rangle$ be arbitrarily fixed. Without loss of generality assume that $\mathbf{x}_0 = \mathbf{0}$ and $\mathbf{x}_{m+1} = \mathbf{1}$. The following holds for the measure μ_3 defined by (2.5) (see also (2.6)):

$$\begin{aligned}\mu_3(\alpha\mathbf{x}; \alpha\mathbf{x}_1, \alpha\mathbf{x}_2, ..., \alpha\mathbf{x}_n) &= \frac{1}{2} + \frac{d_p(\alpha\mathbf{0}, \alpha\mathbf{x}) - d_p(\alpha\mathbf{x}, \alpha\mathbf{1})}{2d_p(\alpha\mathbf{0}, \alpha\mathbf{1})} \\ &= \frac{1}{2} + \frac{\left(\sum_{i=1}^n (\alpha x_i)^p\right)^{1/p} - \left(\sum_{i=1}^n (\alpha - \alpha x_1)^p\right)^{1/p}}{2\left(\sum_{i=1}^n \alpha^p\right)^{1/p}} \\ &= \frac{1}{2} + \frac{\alpha d_p(\mathbf{0}, \mathbf{x}) - \alpha d_p(\mathbf{x}, \mathbf{1})}{2\alpha d_p(\mathbf{0}, \mathbf{1})} = \mu_3(\mathbf{x}; \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n).\end{aligned}$$

The latter justifies the zero-homogeneity of μ_3. A similar property follows for remaining measures μ_i, for $i = 1, 2, 4$. The proof is omitted here. □

Note 1. If the distance $\rho^* = d_{rad}$ (see (2.4)) is used in (2.5) then all measures μ_i (for $i = 1, 2, 3, 4$) are homogeneous functions of the zero order. Condition of the zero order homogeneity is the basic requirement for the measures discussed (independence from the scale – see [17]).

Note 2. The measures may be considered as the utility functions and may be treated as functions of consumption demand for which the zero order homogeneity means that the so-called "no money impact" phenomenon takes place on the market (see [19]).

Property 2. If $\mathbf{x} \in \langle \mathbf{x}_0, \mathbf{x}_{m+1} \rangle$, then equality $\mu_3(x) = \mu_4(x)$ holds if and only if: $\rho(\mathbf{x}_0, \mathbf{x}) = \rho(\mathbf{x}_{m+1}, \mathbf{x})$ or $\rho(\mathbf{x}_0, \mathbf{x}) = 1 - \rho(\mathbf{x}_{m+1}, \mathbf{x})$.

Proof. Let $a = \rho(\mathbf{x}_0, \mathbf{x})$ and $b = \rho(\mathbf{x}_{m+1}, \mathbf{x})$. Based on (2.7) we have $(1/2)(1 + a - b) = a(a + b)^{-1}$ from which it follows $a(a - 1) = b(b - 1)$. The last equality holds if and only if $a = b$ or $a = 1 - b$. □

In the special case when $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{x}_{m+1} = \mathbf{1}$ and $n = 2$, where $\rho(\mathbf{x}, \mathbf{y}) = \rho^*(\mathbf{x}, \mathbf{y})/\sqrt{2}$ and ρ^* denotes Euclidean metric, the equality of measures μ_3 and μ_4 holds only on the diagonals of a unit square.

Property 3. If $\mathbf{a} = (a, a, ..., a) \in \langle \mathbf{0}, \mathbf{1} \rangle$ and ρ in (2.5) coincides with either a radar distance d_{rad} or with Minkowski's metric d_p, $p \geq 1$, then the following holds:

$$\mu_1(\mathbf{a}) = \mu_2(\mathbf{a}) = \mu_3(\mathbf{a}) = \mu_4(\mathbf{a}) = a,$$

where measures μ_i (for $i = 1, 2, 3, 4$) defined by (2.7).

Property 3 implies that given the worst and the best objects have the measures equal to 0 and 1, respectively, then the intermediate object described by the vector $\mathbf{1/2}$ has a measure amounting to 0,5. This property permits to group the objects depending on their respective measures.

The remarks below enable in particular cases to determine the surfaces and indifference domains induced by μ_3 and μ_4.

Note 3. It can be shown that with an Euclidean metric d_2, two objects $\mathbf{x}_i$, $\mathbf{x}_j$ (for $i, j \in \{1, 2, ..., m\}$) have the same measure $\mu_3(\mathbf{x}_i) = \mu_3(\mathbf{x}_j)$ if and only if the difference of the squares of their distances from the best object is equal to the difference of the squares of their distance from the worst object, i.e.:

$$\mu_3(\mathbf{x}_i) = \mu_3(\mathbf{x}_j) \iff d_2^2(\mathbf{x}_i, \mathbf{x}_{m+1}) - d_2^2(\mathbf{x}_j, \mathbf{x}_{m+1}) = d_2^2(\mathbf{x}_i, \mathbf{x}_0) - d_2^2(\mathbf{x}_j, \mathbf{x}_0).$$

From the above it follows that if objects $\mathbf{x}_i$, $\mathbf{x}_j$ are equidistant from the best object $\mathbf{x}_{m+1}$ and from the worst object $\mathbf{x}_0$, i.e.:

$$d_2(\mathbf{x}_i, \mathbf{x}_{m+1}) = d_2(\mathbf{x}_j, \mathbf{x}_{m+1}) \quad \text{and} \quad d_2(\mathbf{x}_i, \mathbf{x}_0) = d_2(\mathbf{x}_j, \mathbf{x}_0),$$

then $\mu_3(\mathbf{x}_i) = \mu_3(\mathbf{x}_j)$. Similarly, for the measure μ_4, if the objects $\mathbf{x}_i$, $\mathbf{x}_j$ are equidistant from $\mathbf{x}_{m+1}$, we have:

$$\rho(\mathbf{x}_i, \mathbf{x}_{m+1}) = \rho(\mathbf{x}_j, \mathbf{x}_{m+1}),$$

then $\mu_4(\mathbf{x}_i) = \mu_4(\mathbf{x}_j)$.

4 Models of measure of concentration

In the case of evaluating concentration of "goods" attributed to a given set of objects, in practice one focuses on how much a vector $\mathbf{x}$ (which coordinates represent the share of goods possessed by each of n objects) differs from the so-called *egalitarian vector* with all coordinates equal to $1/n$ denoted by $\mathbf{e} = (1/n, 1/n, ..., 1/n) \in \mathbb{R}^n$ (see [4, 21]). For further consideration we introduce also the vector $\mathbf{s} = (0, ..., 0, 1) \in \mathbb{R}^n$ corresponding to *an extreme good concentration vector*. Both vectors $\mathbf{e}$ and $\mathbf{s}$ play vital role to test different measures in concentration analysis (see [4, 21]).

We introduce now a special set:

$$\Omega = \{\mathbf{x} = (x_1, ..., x_n) \in [0, 1]^n : x_1 + x_2 + ... + x_n = 1,\ x_i \geq 0,\ i = 1, 2, ..., n\}$$

representing the set of all n-element *structures*.

Let $P : \Omega \to \Omega$ denote an operator designated to order coordinates of any vector according to (2.1). The operator P transforms a vector $\mathbf{x} = (x_1, ..., x_n) \in \Omega$ into a vector $\mathbf{x}' = (x'_1, ..., x'_n) \in \Omega$ by permuting its coordinates so that

$x'_1 \leq x'_2 \leq ... \leq x'_n$. Such operator P is called *an ordering operator* with the associated notation assigned as $\mathbf{x}' = P(\mathbf{x})$. Upon coordinates' re-ordering one can define a function $C : \Omega \to [0,1]^n$ (called *a cumulation operator* – see [4,21]) which maps vector $\mathbf{x}' \in \Omega$ into $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2, ..., \hat{x}_n)$ according to $\hat{x}_i = \sum_{j=1}^{i} x'_j$, for $i = 1, ..., n$.

Obviously, $\hat{\mathbf{x}} = C(\mathbf{x}') = C(P(\mathbf{x}))$, $\mathbf{x}' \in \Omega$. Given a fixed measure of distance d, one can formulate the following concentration measures [4, 6, 21]:

$$\Psi_1(\mathbf{x}) = \frac{d\big(C(\mathbf{e}), C(P(\mathbf{x}))\big)}{d\big(C(\mathbf{e}), \mathbf{s}\big)}, \quad \Psi_2(\mathbf{x}) = \frac{d\big(\mathbf{e}, P(\mathbf{x})\big)}{d\big(\mathbf{e}, \mathbf{s}\big)}, \tag{4.1}$$

where $\mathbf{e}, \mathbf{s} \in \Omega$.

Recall (see [4,6,21]) that if Ψ is a measure of concentration and $f : [0,1] \to [0,1] \subset \mathbb{R}$ is a monotonically increasing surjection then $f(\Psi)$ is also a measure of concentration. In case of measures of concentration economists and sociologists have set requirements imposed on their properties. It should be stressed that in general researches are in agreement that a measure of concentration should, at the very least, have the following properties:

1. the measure is equal to zero if the good is equally distributed among all objects (egalitarian distribution);

2. the values of the indicator are in agreement with the principle of transfers, which states that a transfer from a "poorer" object to a "richer" object of any part of the former good will cause an increase in the inequality within the population;

3. the *transfer sensitivity axiom*: the impact of a transfer from a "rich" object to a "poor" object on the value of the indicator is directly proportional to how "rich" the former object is;

4. the indicator assumes its maximal value when all goods are possessed by a single object.

Note 4. It can be shown that the common indicators like Herfindahl - Hirschman HHI or Gini $GINI$ in their normalized forms (HHI^* and $GINI^*$) are expressible in terms of vectors $\mathbf{e}$ and $\mathbf{s}$ and Minkowski's metric d_p (2.3) according to the formulae (see [7]):

$$\mathrm{HHI}^*(\mathbf{x}) = \left[\frac{d_2\big(\mathbf{e}, P(\mathbf{x})\big)}{d_2(\mathbf{e}, \mathbf{s})}\right]^2, \quad \mathrm{GINI}^*(\mathbf{x}) = \frac{d_1\big(C(\mathbf{e}), C(P(\mathbf{x})\big)}{d_1(C(\mathbf{e}), \mathbf{s})}. \tag{4.2}$$

Combining (4.1) and (4.2) we obtain

$$\mathrm{HHI}^*(\mathbf{x}) = \Psi_2^2(\mathbf{x}) \quad \text{and} \quad GINI^*(\mathbf{x}) = \Psi_1(\mathbf{x}).$$

For $\mathbf{x}, \mathbf{y} \in \Omega$ one can show that functor d_R defined by

$$d_R(\mathbf{x}, \mathbf{y}) = |R^*(\mathbf{x}) - R^*(\mathbf{y})|, \tag{4.3}$$

where $R^*(\mathbf{x})$, $R^*(\mathbf{y})$ denote the ratios of areas of polygons created by radar charts of vectors $\mathbf{x}$, $\mathbf{y}$, respectively, and the area of a polygon induced by the radar chart of vector $\mathbf{x}_{max} = (1, 1, ..., 1)$, i.e. a function defined by:

$$R^*(\mathbf{x}) = R^*(x_1, x_2, ..., x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i x_{i+1} \quad \text{with} \quad x_{n+1} = x_1.$$

Indicator d_R defines the distance between structures $\mathbf{x}$ and $\mathbf{y}$.

Note 5. If in (4.1) defining the indicator $\Psi_1(\mathbf{x})$ the functor d_R is applied then the so-called radar indicator of concentration is obtained

$$\Psi_{rad}(\mathbf{x}) = d_R\big(C(\mathbf{e}), C(P(\mathbf{x}))\big)/d_R\big(C(\mathbf{e}), \mathbf{s}\big).$$

Note 6. In the case of measuring concentration not every distance function listed in (4.1) meets four requirements specified above by the practitioners. However, as shown in [4, 21] the measures constructed by using Minkowski's metric and three other distances based on graphical interpretations in the form of radar charts (especially distance d_R defined by (4.3)) specify the above four constraints.

5 Results and empirical research

To confirm the need of using multiple measures simultaneously for ordering objects characterized by n-features and to evaluate their differentiation (e.g. of wages) we present an application of the above analysis on some real empirical data[1].

The analysis and experiments are performed on the data acquired from 10 branches of a certain retail bank. For each of them we have collected information about personnel costs of employees (see Table 2) from their central Human Resources (denoted as HR) system as well as indicators rating the efficiency of these branches (see Table 1): $X1$ - ROA, $X2$ - ROE, $X3$ cost income ratio (operational costs in income from their core business), $X4$ - ratio of clients' deposits in balance sheet total/assets, $X5$ capital adequacy ratio, $X6$ - ratio of performing credits to total credits. The respective values of these indicators upon normalisation:

$$\frac{x_i - \min_{\{i\}}(x_i)}{\max_{\{i\}}(x_i) - \min_{\{i\}}(x_i)}$$

are listed in Table 1. Columns $W1$ and $W3$ contain two measures representing distances from a negative object - in this case the negative object being $\mathbf{x}_0 = (0, 0, ..., 0)$. We used here a measure defined by (2.7) and a metric defined by (2.3) for $p = 1$ and $p = 4$. Additionally, the column $W2$ contains an indicator defined by (2.7) resorting to a metric (2.3) with $p = 2$. Rankings of branches as per these indicators are present in columns $R1$, $R2$, $R3$, respectively. Rankings $R1$ and $R2$ are identical but widely different from $R3$. The last column ($R1$-$R3$) presents changes in positions of branches between rankings $R1$ and $R3$.

[1] Source - own research.

This simple juxtaposition shows that a choice of metric can greatly influence the position of an object. However, that might not be always the case.

Ranking with respect to indicator $W2$ shows similar result as in case of indicator $W1$, which is constructed by using a different formula and the same metric. This example confirms our understanding of indicators' properties and the choice of one of them for analysis of empirical data.

This is crucial especially in a capitalist economy. If the bonus pool depends on this ranking then the choice of a measure becomes a "sensitive" situation.

Table 1. Values of chosen indicators after normalisation.

	X1	X2	X3	X4	X5	X6	W1	W2	W3	R1	R2	R3	R1-R3
B01	0.556	0.545	0.077	0.992	0.000	0.996	0.528	0.517	0.772	7	7	3	4
B02	0.333	0.773	0.308	1.000	0.625	0.266	0.551	0.539	0.711	3	3	4	-1
B03	0.111	1.000	0.538	0.888	1.000	1.000	0.756	0.667	0.887	1	1	1	0
B04	0.444	0.273	0.769	0.875	0.750	0.158	0.545	0.535	0.682	4	4	6	-2
B05	1.000	0.636	1.000	0.197	0.250	0.114	0.533	0.521	0.775	6	6	2	4
B06	0.778	0.455	0.646	0.756	0.625	0.614	0.646	0.639	0.669	2	2	7	-5
B07	0.222	0.182	0.846	0.308	0.875	0.769	0.534	0.525	0.702	5	5	5	0
B08	0.000	0.318	0.385	0.058	0.938	0.051	0.292	0.358	0.605	10	10	9	-1
B09	0.667	0.614	0.231	0.000	0.813	0.000	0.388	0.422	0.600	8	8	10	-2
B10	0.889	0.000	0.000	0.320	0.375	0.727	0.385	0.422	0.628	9	9	8	1

We calculated the differentiation of personnel costs for 10 branches of some retail bank. Table 2 contains personnel costs assigned to each of employees $(Q1, ..., Q10)$ and values of coefficients of concentration HHI^*, $GINI^*$ and $RADAR$ (see notes 4 and 5) given to the third decimal place. For further illustration we also include values of indicator $C3$ (the ratio of three largest personnel costs assigned to employees to the total personnel cost of the branches), the mean (average personnel cost per employee) and the value of the volatility index V (ratio of standard deviation to the mean). Table 2 shows that when observing the values of a popular $GINI^*$ indicator (or its normalised values, to be precise) one cannot observe a significant concentration difference of income in branch $B4$ as compared to $B5$–$B10$. A similar situation occurs for HHI^* and branches $B1$, $B2$, $B3$. Moreover, Table 2 contains in its last five rows $(R1, ..., R5)$ the rankings as defined by coefficients of concentration V, $C3$, HHI^*, $GINI^*$, $RADAR$. It is clear that the three coefficients of concentration (HHI^*, $GINI^*$, $RADAR$) result in different ordering of branches by levels of concentration of personnel costs. The latter coincides with our intuition as each of these coefficients is "sensitive" to a different aspect of changes in the structure of costs of a branch in relation to an egalitarian structure $\mathbf{e}$ (see equations (4.1), (4.2) and (4.3)).

The techniques presented in this work for constructing indicators utilizing distance from model objects allow analysts to create measures that are directed

Table 2. Values of personnel costs in 10 branches of a retail bank.

	B01	B02	B03	B04	B05	B06	B07	B08	B09	B10
Q1	6	7.8	7.7	11.189	9.741	5.2	10.583	8.5	8.328	12.308
Q2	6.5	8.3	8.2	11.189	9.741	7.964	10.583	9.343	8.328	12.508
Q3	7	8.8	8.7	11.189	9.741	10.728	10.583	10.269	8.328	12.708
Q4	8	9.8	9.7	11.189	9.741	13.492	10.583	11.287	8.328	12.908
Q5	8.5	10.3	10.2	11.189	9.741	16.256	10.583	12.406	8.328	13.108
Q6	9	10.8	10.75	11.189	9.741	19.02	10.583	13.636	8.328	13.308
Q7	12	12	11.95	11.189	9.741	21.784	20.817	14.988	16.338	13.508
Q8	20	12	12.3	11.189	28.765	24.548	23.419	16.474	17.505	15.577
Q9	30	22.9	23.3	11.189	28.765	27.312	24.72	18.108	18.672	21.807
Q10	50	54.3	54.2	56.299	28.765	30.076	38.043	41.989	30.809	56.076
Mean	15.700	15.700	15.700	15.700	15.448	17.638	17.050	15.700	13.329	18.382
V	0.860	0.860	0.861	0.862	0.654	0.450	0.526	0.589	0.536	0.699
C3	0.637	0.586	0.572	0.501	0.559	0.465	0.505	0.488	0.503	0.508
HHI^*	0.082	0.082	0.082	0.083	0.035	0.023	0.031	0.039	0.032	0.054
$GINI^*$	0.452	0.385	0.389	0.287	0.287	0.287	0.287	0.287	0.287	0.287
$RADAR$	0.566	0.520	0.524	0.432	0.379	0.344	0.383	0.394	0.386	0.417
R1	3.5	3.5	2	1	7	10	9	6	8	5
R2	1	3	2	8	4	10	6	9	7	5
R3	3	3	3	1	7	10	9	6	8	5
R4	1	3	2	2	7	7	7	7	7	7
R5	1	3	2	4	9	10	8	6	7	5

at tracing changes deemed most important by management. The above, relatively simple, example shows that a rating of a branches efficiency is sensitive to measuring the distance within the model. Hence, it seems helpful to include in managerial briefs grading branches efforts not one but multiple synthetic measures used to order/rate objects defined by multiple specific indicators. Similarly, HR reports tracing changes in personnel costs structures (or just salaries) in individual units should contain not one but many other coefficients of concentration.

6 Conclusions

This work formulates the necessary conditions to be fulfilled while ordering objects with n features. In literature ordering of such objects is usually done via a certain non-directly observable variable(s), most often called a synthetic

measure, development measure or concentration coefficient. Our studies show that most of proposed indicators can be expressed as functions of distance from one or two model objects. We indicated here that synthetic indicator W based on calculating a distance from models should be defined in an interval $[0, 1]$. Moreover, the distances between any two objects $\mathbf{x}, \mathbf{y} \in X$, have to fulfill at least two conditions: (1) $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$ and (2) $d(\mathbf{x}, \mathbf{x}) = 0$.

Universal properties of indicators also include invariability with respect to changes in scale when all variables have values on an interval scale (or only on a quotient scale). But the indicator must be sensitive to any other variation of the object features and such changes must result in changing the value of indicator. This paper shows that for measuring the concentration not every distance results in a measure with desired properties. However, as proved in [4] the measures constructed by Minkowski's metric and three distances based on graphical interpretations in the form of radar charts (especially distance d_R defined by (4.3)) meet all the properties specified in Section 4.

References

[1] H.R. Anderberg. *Cluster Analysis for Applications*. Academic Press, New York, 1973.

[2] J.P. Barthelemy, F. Brucker and C. Osswald. Combinatorial optimization and hierarchical classifications. *Quarterly Journal of Operations Research*, **2**(3):179–219, 2004. https://doi.org/10.1007/s10288-004-0051-9.

[3] Z. Binderman, B. Borkowski, G. Koszela, R. Kozera and W. Szczesny. The choice of synthetic measures to assessment economic effects. *Quantitative Methods in Economics*, **18**(1):7–17, 2017.

[4] Z. Binderman, B. Borkowski, A. Prokopenya and W. Szczesny. Applications of dissimilarity measures of objects ordering and concentrations. *Computer Algebra Systems in Teaching and Research*, **5**(1):23–39, 2015.

[5] Z. Binderman, B. Borkowski, Y. Shachmurove and W. Szczesny. An application of radar charts to geometrical measures of structures' of conformability. *Quantitative Methods in Economics*, **12**(1):1–13, 2011.

[6] Z. Binderman, B. Borkowski and W. Szczesny. Radar coefficient of concentration. *Quantitative Methods in Economics*, **13**(2):7–21, 2012.

[7] Z. Binderman, B. Borkowski and W. Szczesny. Applications of Minkowski's metric in measuring changes of concentration of value added in agriculture. *Acta Scientiarum Polonorum. Oeconomia*, **14**(1):17–28, 2015.

[8] L. Breiman. Bagging predictors. *Machine Learning*, **24**(2):123–140, 1996. https://doi.org/10.1007/BF00058655.

[9] L. Breiman. Arcing classifiers. *The Annals of Statistics*, **26**(3):801–849, 1998.

[10] L. Breiman. Bias-variance, regularization, instability and stabilization. In C. M. Bishop(Ed.), *Neural Networks and Machine Learning*, pp. 27–56, Berlin, Heiderlberg, New York, 1998. Springer-Verlag.

[11] A.J. Cole. *Numerical Taxonomy*. Academic Press, New York, 1969.

[12] R.M. Cormack. A review of classification. *Journal of the Royal Statistical Society Series A (General)*, **134**(3):321–353, 1971. https://doi.org/10.2307/2344237.

[13] D. Davies and D. Boulding. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1**(2):224–227, 1979. https://doi.org/10.1109/TPAMI.1979.4766909.

[14] J.L. Doob. *Measure theory*. Springer-Verlag, New York, 1994.

[15] A.D. Gordon. *Classification*. Chapman and Hall, London, 1981.

[16] C.L. Hwang and K. Yoon. *Multiple attribute decision making: methods and applications*. Springer-Verlag, New York, 1981.

[17] D.H. Jackson. *The stability of classification of binary attribute data. Technical Report 70-65*. Cornell University, 1970.

[18] I.T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, 1986.

[19] E. Panek. *Ekonomia Matematyczna*. Wydaw. AE, Poznań, 2000. (in Polish)

[20] J.H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, **58**(301):236–244, 1963. https://doi.org/10.1080/01621459.1963.10500845.

[21] Z.J. Zawistowski, W. Szczesny, B. Borkowski, R. Kozera and Y. Shachmurove. Alternative method of measuring concentration. *Applied Mathematics and Information Sciences*, **10**(1):11–19, 2016. https://doi.org/10.18576/amis/100102.