Recent Advances in Business Analytics. Selected papers of the 2021
KNOWCON-NSAIS workshop on Business Analytics pp. 53–61
DOI: 10.15439/2021B3
ISSN 2300-5963 ACSIS, Vol. 29

Interval-valued semantic differential in multiple criteria and multi-expert evaluation context: possible benefits and application areas

Jana Stoklasová
School of Business and Management,
LUT University Yliopistonkatu 34,
53851 Lappeenranta, Finland
Email: jana.stoklasova@lut.fi

Abstract—**The paper discusses the possibilities of adapting the recently introduced interval-valued semantic differential method to the multiple-criteria decision-making and evaluation context. It focuses on the differences and common ground of the intended use of the original semantic differentiation method and general multiple-criteria evaluation problems. The paper identifies the aspects of the interval-valued modification of the method that can be useful in multiple-criteria evaluation and also aspects that can be beneficial in the multi-expert evaluation setting and also possible limitations stemming from the transition to the multiple-criteria (or multi-expert) evaluation context. Finally the paper suggests potential application areas for the (interval-valued) semantic differential based methods.**

I. Introduction

THE set of methods available for multiple-criteria and multi-expert evaluation problems is large and it is being continuously expanded (see e.g. [1], [2], [3], [4], [5], [6]). The currently available methods include methods for weights determination (see e.g. [7], [8], [9]), methods for the standardization of values of criteria, various methods for the aggregation of values across different criteria [10], [11], [12], [13], methods for preference representation [14], [15], [16] and aggregation [17], [18], [19]. We have specific methods based on pairwise comparisons (see e.g. [20], [21], [22]), methods utilizing ideals in the evaluation process [23], [24], special methods for ordinal data [25], [26], [27], methods equipped to deal with different types of uncertainty [28], [29], [30], [31], [32], [33], [34], methods capable of dealing with linguistic inputs/outputs and to process natural language [35], [36], [37], [38], methods for consensus modeling and analysis [39], [40], [27], [41]. The list is definitely not complete, nor is it reasonably structured. The point we would like to make here is that currently there are many methods available to model and assist with human-like decision making. They focus on different aspects of the evaluation in these problems and are able to reflect many different specific features of the decision-makers, of the alternatives, of the scales used for the evaluation etc. The behavioral perspective has entered the multiple-criteria and multi-expert evaluation and decision-making domain long ago and is still gaining momentum [42], [43], [44], [45], [46], [47], [48], [49], [50]. Most of the methods assume at least some sort of measurability of the features of the alternatives that are being evaluated, or of the values of the criteria that are being used in the process; some circumvent the requirement of measurability by pairwise comparisons, by the use of linguistic assessments, by the use of ordinal values only etc. There are, however, very few methods in the multiple-criteria and multi-expert evaluation field that would be focused or tailored for dealing with intangible criteria.

The research was supported by LUT research platform AMBI - Analytics-based management for business and manufacturing industry.

Even though current research is aiming also on the ability of computers and models to recognize, process, mimic and interpret emotions [51], the efforts to incorporate affects and other less tangible criteria in evaluation models are limited. This might be stemming from the difficulties with measuring or obtaining the information on affect and other less tangible concepts like attitudes, political preferences, religion, values, connotative meaning of words etc. On the other hand there are methods in psychology, anthropology, linguistics and related fields that are designed for the very purpose of capturing non-measurable and intangible concepts. One of these methods, the semantic differential method by Osgood, Suci and Tannenbaum [52] is going to be investigated in this paper. We will describe the main principles of this method, briefly recall its recent interval-valued generalization [53], [54] and identify how the concepts intended for the capturing of intangible characteristics can be applied in the multiple-criteria evaluation and multi-expert evaluation setting. We will particularly focus on those aspect that are crucial in the original definition of this tool and have psychological value (such as partial projectivity, the requirement on the bipolar adjectives scales being non-descriptive for the evaluated alternative/concept, etc.) and their meaningfulness, usefulness or potential drawbacks if transferred directly into the multiple-criteria evaluation setting. Our aim is to identify those features of the semantic differentiation method (and its interval-valued generalization) rooted in its original social-science use, that can be beneficial in multiple-criteria evaluation models.

II. Semantic differential and its main features

Semantic differential (SD) is a method introduced by Osgood, Suci and Tannenbaum in 1957 [52] as a technique for the quantification or representation of connotative meaning of words. Soon enough it found its way to anthropology [55] and obviously also to psychology for the measurement (quantification) of attitudes [56], [57].

The basic tool in the method are bipolar adjective scales that are used for the assessment of the given object/term/concept (the bipolar adjective scales will also be called items in the text for more simplicity). These scales are the basic "measurement" instrument in the method. The scales are assumed to share the same universe, let us say $[a, b] \subset \mathbb{R}$. Some authors suggest that $0 \in [a, b]$, some suggest that $a > 0$, some that $a = -b$, but the actual form of the scale influences mainly the comfort and reliability of the respondent's answer. Let us now assume that the underlying scale is a continuum with extremes a and b representing the opposite poles of the scale. Originally, discrete (7-point) scales were used in [52] but the actual form of the scale was more tailored for that time's methods of data collection and analysis. The transition to continuous scales is of no actual consequence for the design and performance of the semantic differential method. In other words we can also use continuous scales instead of discrete ones and the method works as well.

The method targets the less tangible aspects of the evaluated object/concept, that is, it intends to capture the connotative (individual-specific) meaning of the concept, reflect the individual's experience and specifics. In social psychology the ability to capture not-measurable aspects connected with the assessed concept led to the use of semantic differential in the quantification of attitudes (mainly in the three-factor model of attitudes where attitudes are assumed to have cognitive, conative and affective components - the latter two being difficult to directly measure). It is therefore suggested by Osgood et al. ([52]) to avoid such bipolar-adjective scales that would have actual descriptive power over the evaluated object. Note that this is a very particular requirement for a method that should be considered for multiple-criteria evaluation. There are, however, good psychological reasons behind this requirement. First the use of descriptive items (e.g. sharp-blunt for the description of a knife) and non-descriptive items (e.g. happy-sad for the description of the same knife) together in one assessment tool (inventory or set of bipolar scales) could result in a lower reliability of the non-descriptive items. The respondents might simply wonder whether they understand the evaluation task well as some items have clear connection with the evaluated concept while others do not. Second the use of descriptive items provides a description of the object/concept rather than its actual evaluation. Third the use of non-descriptive scales decreases respondents' ability to provide desirable, "fake-good" or "fake-bad" answers, which decreases the potential deliberate distortion of information by the respondent/evaluator. The whole procedure of using a semantic differential in the assessment of connotative meanings of concepts or the assessment of attitudes of the respondents towards these concepts can be summarized in the following steps (more details can be found in [52]):

1) Generate a set $S = \{s_1, \ldots, s_n\}$ of bipolar-adjective scales. It should contain sufficiently many scales, the meanings of their endpoints should be understandable to the potential evaluators (respondents), enough of these scales should be non-descriptive for the concepts to be evaluated.
2) Carry out pilot run where all these scales are used to assess some concepts by a representative sample of the target population.
3) Carry out a factor analysis (both exploratory and confirmatory versions are suggested) to determine whether the dimensionality of the original set of scales can be reduced to just a few underlying factors. The factors are to be identified ideally in such a way that they could be named and interpreted accordingly (apply factor notation if needed). For example in [52] three factors were identified: Evaluation, Potency and Activity. These factors are expected to represent orthogonal evaluation dimensions. Let us assume that k factors $F_1, \ldots, F_k$ are identified. Then the factor loadings of the scale $s_i \in S$ for factors $F_1, \ldots F_k$ can be denoted $f_{1_{s_i}}, \ldots, f_{k_{s_i}}$ respectively.
   Note that the factor loadings (and the factors) are therefore domain- and culture-specific. In other words the factor analysis should be performed every time we apply the chosen scales to the evaluation/assessment of concepts in a different context, also when we change the target population. Given the fact that the extreme values (poles) of the scales are described linguistically, every language mutation of the scales should have its own factor analysis performed.
4) Select a subset of the bipolar-adjective scales $Z \subseteq S$, $Z = \{z_1, \ldots, z_m\}$ that would be used for the given application. Usually scales that sufficiently load at least one factor are used, it is also good to use scales that would allow all the factors to be "measured" and also to allow for repeated measurement of each factor.
5) Obtain data from the respondents. In other words let each respondent assess the concept using all m chosen bipolar-adjective scales $z_1, \ldots, z_m$. If the assessment of the concept/object by a respondent X on scale z_i is denoted as x_{z_i}, $i = 1, \ldots, m$, then the object/concept O is represented as a point O^X in the k-dimensional space $[a, b]^k$ with the following coordinates:

$$O^X = \left(\frac{\sum_{i=1}^m x_{z_i} \cdot f_{1_{z_i}}}{\sum_{i=1}^m |f_{1_{z_i}}|}, \ldots, \frac{\sum_{i=1}^m x_{z_i} \cdot f_{k_{z_i}}}{\sum_{i=1}^m |f_{k_{z_i}}|} \right) = (x_{F_1}, \ldots, x_{F_k}). \quad (1)$$

   In other words the coordinates are the factor-loading weighted average of the answers provided by the respondent. Sometimes only the contribution of the item to the factor with the highest factor loading is reflected in the practical applications of semantic differential.

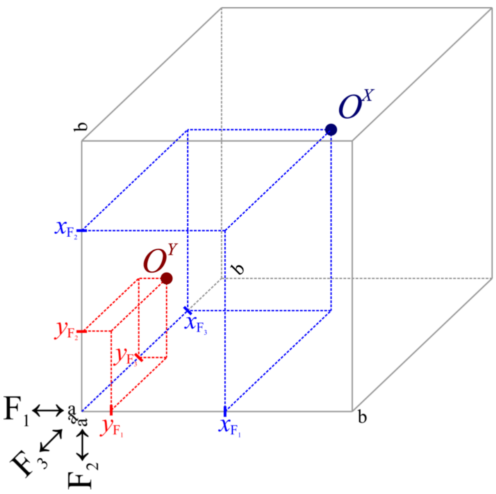

Fig. 1. An example of the output of the standard semantic differential method [52] for two objects/concepts X and Y. Three factors are assumed. The assessment of X is represented by the point $O^X = (x_{F_1}, x_{F_2}, x_{F_3})$ in the three-dimensional space defined by the factors F_1, F_2 and F_3, the assessment of Y is represented by the point $O^Y = (y_{F_1}, y_{F_2}, y_{F_3})$.

The above described procedure allows for the representation of a connotative meaning of a concept or an attitude towards that concept to be represented as a point in an k-dimensional space where each dimension represents one factor (higher level characteristic of the object/concept) that is orthogonal to the other factors. An example of the output of the standard semantic differential method is presented in Fig. 1.

To summarize, the benefits of the method as proposed by Osgood et al. ([52]), assessed from a multiple-criteria evaluation perspective, are the following:

- The factor analysis applied provides a few orthogonal evaluation dimensions to work with. This means that a visualization of the results, that is easy to understand, might be possible.
- The use of non-descriptive bipolar-adjective scales prevents deliberate distortions of the data by respondents.
- The use of bipolar-adjective scales provides a "projective-like" feature of the data collection that in terms allows for the assessment of less tangible criteria/aspects of the concept.
- The fact that more items have non-zero loadings to the same factor means that we have repeated assessment of each factor.
- Data input using the semantic differential scales is rather simple.
- It is possible to define distances in the $[a, b]^k$ space to decide which representations of objects/concepts are close to each other, which are far from each other.
- As long as the factors are defined with appropriate labels and can be seen as consistent characteristics "measured" by multiple items (repeated "measurement"), the coordinates of the concepts in the $[a, b]^k$ space can be interpreted. It is also possible to define "desired" or "undesired" values in this space, that is to define ideals to be used in the evaluation or decision-making.
- There is no need for aggregation across the factors. Aggregation within one factor (1) is understood as repeated measurement of the factor, other aggregation is not necessary. The final representation of the result of semantic differentiation can therefore be understood as virtually lossless. The aggregation across factors, if needed, can also be done, for example, via the definition of the distance from a given ideal in the $[a, b]^k$ space.

It is therefore clear that many features of the semantic differential can be seen as beneficial for the standard multiple-criteria or multi-expert evaluation. On the other hand there are certain clear limitations or drawbacks of the method when considered for practical multiple-criteria evaluation:

- The factor analyses need to be done. As the factors, their number, definition and loadings of the scales can be context and culture dependent, it might take a lot of time to set up the scales and find their factor loadings.
- Also a conversion to other languages and other domains of application requires new factor analyses. The language issues are even more complex than might be apparent at first sight. If the tool is calibrated, for example, for English scales for a given context of application (factor loadings of items are determined with English labels of the endpoints of the bipolar adjective scales) it should still not be directly applied with non-native speakers of English, unless these were present in the original sample used to determine the factors and their loadings.
- The factors are stemming from the factor analysis. They are therefore constructed and might not have clear interpretation. This could limit the interpretability of the results of semantic differential in evaluation applications.
- The issue of concept-scale interaction and lower perceived scale relevance may be present [58]. This means that the respondents might see some scales as inappropriate for the assessment of a given concept and thus the value of the given item provided by them can be arbitrary without the researcher knowing so.
- The method has no means of incorporating uncertainty stemming from lower perceived item relevance for the evaluation of the given object/concept, from the misinterpretation of the meanings of the endpoints of the scales or simply from the inability of the respondent to provide answers using some items because their connection with the assessment might be too value or unclear.
- The single-point in $[a, b]^k$ space might appear much more precise than it should.
- It might not be clear if a "middle" answer means the inability of the respondent to use the given bipolar-adjective scale, or whether his/her assessment is really neutral.

Even though there are clear benefits that speak in favor of the semantic differential being used in multiple-criteria evaluation, there are still some shortcomings that make its use problematic. Some of these shortcomings can be overcome by generalizing the semantic differential into an interval-valued method as proposed by Stoklasa et al. [54]. The interval-valued methods are being applied in other areas as well [33].

III. Interval-valued generalization of the semantic differential

The generalized semantic differential (GSD) method was introduced in 2019 by Stoklasa, Talášek and Stoklasová with the intention of introducing means for the reflection of uncertainty of the answers provided by the respondents in the form of the x_{z_i} values [54]. GSD assumes that each bipolar adjective scale $z_i \in Z$ is accompanied by another scale r_{z_i} designed to assess the relevance of the scale z_i for the assessment of the given object/concept as perceived by the decision maker. The authors suggest a $[0\%, 100\%]$ universe for each relevance scale r_{z_i} for any $i = 1, \ldots, m$. The term "perceived relevance" can be replaced by any potential source of uncertainty of the values x_{z_i} provided by the respondent/evaluator. The source of uncertainty discussed specifically in [54] is the incompatibility (partial or full) of the bipolar adjective scale with the assessment/evaluation task perceived by the respondent. In other words if the scale is perceived as partially irrelevant by the person who is supposed to use it to assess the given concept, the actual value x_{z_i} is not reliable and should not be considered precise. Due to the partial irrelevance of the scale z_i, the value x_{z_i} might be misspecified by the respondent due to the fact that it was difficult to him/her to established a connection between the evaluated object/concept and the bipolar adjective scale. As such the actually expressed value x_{z_i} is in these cases accompanied by an interval of "also possible values" $I_{z_i} = [x_{z_i}^L, x_{z_i}^R] \subseteq [a, b]$, whose length is proportional to the perceived irrelevance of the scale. Stoklasa et al. [54] suggest the use of Dombi's kappa function [59] to parameterize the calculation of the length of the "interval of also possible values" from the the perceived (ir)relevance of the scale, in other words $\kappa(r_{z_i}) = \left|[x_{z_i}^L, x_{z_i}^R]\right|$. This interval is centered around x_{z_i}, if possible. If this is not possible, then it is shifted so that the shift is minimal, the whole "interval of also possible values of x_{z_i} fits within the $[a, b]$ universe and its length calculated using the kappa function is preserved. The final representation of the output of GSD for an object/concept X is the point O^X in the $[a, b]^k$ space determined from the x_{z_i} values by (1) accompanied by the box of uncertainty B^X (or box of also possible values) surrounding it determined by (2), which is a direct analogy to (1) using interval algebra.

$$B^X = \left(\frac{\sum_{i=1}^{m} I_{z_i} \cdot f_{1_{z_i}}}{\sum_{i=1}^{m} |f_{1_{z_i}}|}, \ldots, \frac{\sum_{i=1}^{m} I_{z_i} \cdot f_{k_{z_i}}}{\sum_{i=1}^{m} |f_{k_{z_i}}|} \right) \quad (2)$$

Interval algebra (see [31, p. 103] for more details) is applied to obtain the final outputs from the generalized semantic differential. This method provides not only the outputs available in the original version of the method - that is the representation of the object/concept X as a point O^X in the k-dimensional Cartesian space - but also a box of uncertainty B^X surrounding the point O^X. See Fig 2 for an example of the outputs of the GSD method. The size of the box of uncertainty is proportional (with respect to the selection of parameters for the kappa function) to the average perceived irrelevance of the bipolar adjective scales used for the assessment of the object/concept. In Fig. 2 it is apparent that the items with high factor loadings for the factor F_1 are perceived as much less relevant for the assessment of X than they are for the assessment of Y.

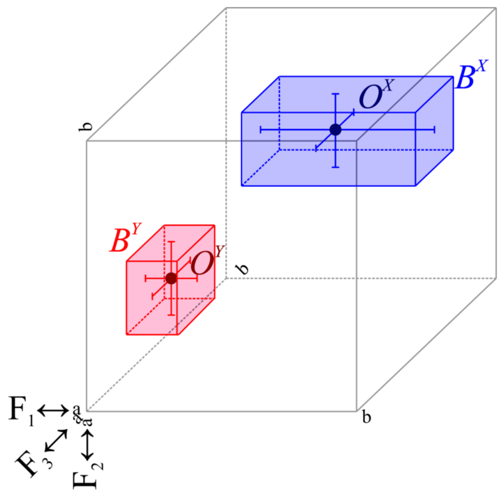

Fig. 2. An example of the output of the generalized semantic differential method [54] for two objects/concepts X and Y. Three factors are assumed. The same two objects X and Y are considered as in Fig. 1 with the same coordinates of O^X and O^Y respectively. Boxes of uncertainty stemming from lower perceived relevance of some scales for the assessment of X and Y are depicted as B^X and B^Y around the O^x and O^Y respectively.

The steps needed to apply GSD are similar to those for SD just with a few minor changes:

1) We again need to have the set $S = \{s_1, \ldots, s_n\}$ of bipolar-adjective scales generated with the same requirements as in SD.
2) We need to administer all those scales to a representative sample of the target population to be able to derive the factors and the factor loadings of the scales (again exploratory and confirmatory factor analysis is recommended). Note that at this point the perceived relevance scales are not used yet. This means that the factors and the factor loadings of the scales are determined independently of the perceived relevance. This also means that if factors and factor loadings are already available for an applicable set of bipolar adjective scales derived for a compatible area of application, these can be used in GSD.
3) We select a subset of the bipolar-adjective scales $Z \subseteq S$, $Z = \{z_1, \ldots, z_m\}$ that will be used for the given application in the same way as for SD. To each of these scales we attach a "perceived relevance" scale r_{z_i}. The r_{z_i} scales, $i = 1, \ldots, m$, are used to capture uncertainty of the x_{z_i} evaluations provided by the respondents. It is possible to label this scale so that it captures different sources of uncertainty of the evaluations as well.

4) We obtain data from the respondents. The values x_{z_i} are assigned "intervals of also possible values" that reflect the irrelevance of the scales and a possible uncertainty of the evaluations stemming from the scale irrelevances (either directly, or through the kappa function).
5) The final representation of the objects/concepts is represented by O^X calculated using equation (1) and by the "box of uncertainty" B^X calculated using the equation (2). Both these representations are depicted graphically (see e.g. Fig. 2).

Allowing the uncertainty in the semantic differentiation process takes care of some of the issues connected with scale-concept interactions. The generalized semantic differential has the same advantages as the original method, plus the ability to reflect uncertainty of the answers provided by the respondents. It can show that the respondent was not very sure about the answers (contributing to particular factors or to all of the factors) by increasing the respective dimension of the "box of uncertainty". As for the disadvantages, the need to perform the factor analyses to get the factors and factor loadings of the bipolar-adjectives scales is still there. Also the factors are defined automatically in the process. All the other limitations or drawbacks listed for SD are mitigated or removed. The method is now slightly more tedious form as it needs to include two sets of scales, meaning a slightly larger workload for the respondents. Also there are more parameters in the GSD to set (the parameters of the kappa function, the framing of the "relevance" scale). Nevertheless, most of the drawbacks listed for the original method can be mitigated by the use of GSD and the method still retains the ability to deal with less tangible criteria. Let us therefore now see, how applicable the method might be in a multiple-criteria or multi-expert evaluation setting.

IV. Generalized Semantic differential and multiple-criteria evaluation

Before we are able to assess the potential benefits of applying GSD in multiple-criteria and multi-expert evaluation, and to suggest the needed modifications of the GSD method for this purpose, we need to define the multiple-criteria evaluation problem first. In multiple-criteria evaluation we assume that we have several objects/alternatives that need to be assessed and assigned a final evaluation of some sort. Usually the expected form of the evaluation is a numerical or vector one, that is in many evaluation methods we are looking for a real-value (or a vector of real values) that would summarize the qualities and the downsides of the alternative sufficiently. We also assume that the k criteria that represent the relevant features of the alternatives are known in advance (usually along with their types, underlying scales and also relative importances). The ultimate goal of the evaluation is then to a) obtain an ordering of the alternatives to be able to decide which ones to select (relative-type evaluation) or b) decide about the acceptability/unacceptability of the alternative (absolute-type evaluation). Let us now have a look at the features of the GSD method and comment on their usefulness or the need for the modification of these aspects for GSD to become a valid multiple-criteria evaluation method.

We need to start with one clear incompatibility between SD or GSD and the multiple-criteria evaluation setting. This is the fact that in GSD (and SD) the factors are defined through factor analysis and thus independent on the user of the evaluation. On the other hand in multiple-criteria evaluation, criteria are usually given and need to be used as defined by the user of the analysis. As we usually expect the k criteria to be independent, we can easily assume that each criterion would be represented by one axis in a k-dimensional Cartesian space. This would mean that if we substitute criteria for factors, we can obtain a method applicable to multiple-criteria evaluation. We can even assume that each criterion is "measured" or assessed repeatedly either through subcriteria, or through different items in a questionnaire or scorecard. Discarding the bipolar adjective scales completely we, however, lose the "projectivity" of the GSD and also the ability to capture less tangible and intangible aspects of the alternatives, as long as we do not have specific items for them in the data input tool (survey, scorecard, etc.). There is always a possibility to keep those bipolar adjective scales that measure the intangible factor(s) that might be relevant for our analysis (e.g. affect) and use externally defined criteria as other dimensions in the final output space. Being able to include the criteria as separate dimensions in the final output space, we can now analyze the other features of the GSD method:

- repeated measurement - semantic differential is built on the idea of repeated measurement of the factors. This is stemming from the fact that factor analysis is applied as a dimensionality reduction technique here. It also means that GSD is ready to process e.g. data from questionnaires where criteria are being assessed by more items. The aggregation of the values provided via different items in a questionnaire or a scorecard can be done either by arithmetic mean or any other feasible aggregation operator, weights can also be reflected, if needed (but weights of items in a questionnaire might not be frequently available).
- ability to capture less tangible aspects/criteria - if bipolar adjectives scale that are not descriptive for the evaluated object are kept in the pool of the items and their respective factor(s) constitute(s) separate dimension(s) in the final output space, then this feature is maintained. On the other hand factor analysis needs to precede the use of the bipolar adjective scales to know which ones are contributing to the desired factor.
- if the intangible criteria/aspects are not important in the evaluation process, then factor analysis might not be needed and the method is much simpler to apply as it does not longer require pre-analysis and an availability of a sample prior to the main analysis/evaluation.
- simple data input procedure - the data input can still be done through questionnaires, where groups of items would contribute to particular criteria. Each item can also

be assigned a scale for the reflection of uncertainty of the answer provided through this scale. Instead of "scale relevance", it might be better to talk about evaluator's confidence with the answer or something similar, though. Afterwards the kappa function can again be used to calibrate the method for the given purpose and to calculate the length of the interval of also possible values based on the confidence with the particular answer.

- graphical outputs - the original SD method was frequently shown to result in three factors. This allowed for a simple three-dimensional graphical representation if the outputs as points in the three-dimensional Cartesian space (called semantic space). For more factors or criteria, or simply for more dimensions of the output space, graphical outputs might not be achievable or easily understood. Still the intuition from three dimensional graphical summaries of the outputs (e.g. those presented in Fig. 2 and Fig. 3) can prove useful in explaining how the methods works in higher dimensions of the output space.
- presentation of results without final aggregation - this is one of the very desirable properties for multiple-criteria but also multi-expert evaluation. The design of the outputs allows for separate treatment of all the criteria representing the dimensions in the output space. The output objects (points or boxes of uncertainty around them) can be defined without an explicit knowledge of the relative importances of the criteria. If the evaluations represent outputs for different experts, the weights of experts do not need to be known either.

If a final ordering of the alternatives is required, one needs to be able to aggregate the information across all the dimensions of the output space. For this we can either introduce weights of criteria, or simply work with the k-dimensional representations of the objects directly and define distances on them. The multiple-criteria evaluation setting has the benefit of being able to define the most desired values of the criteria, and based on them the ideal (potentially non-existent) alternative, or at least the evaluation thereof (see the preference directions in Fig. 3 and the "ideal" evaluation defined based on them). The evaluation task can then be approached by defining distances between the evaluations of the alternatives (up to k-dimensional entities) from the ideal evaluation (also more ideals can be considered like e.g. in TOPSIS). The introduction of uncertainty in the SD represented by GSD then allows for the determination of interval-valued distances (for example shortest distance from the box of uncertainty to the ideal and longest distance from the box of uncertainty to the ideal defining the interval of possible distances - see Fig. 3). Overall GSD has the needed properties to be applied in multiple-criteria evaluation:

- it can handle multiple criteria (including less tangible ones - see the discussion above)
- it is capable of handling uncertainty of the evaluations with respect to the (sub)criteria
- the uncertainty can be assessed using a simple questionnaire-based input procedure. It does not require the respondent to be able to express uncertainty/risk in a complex way and can still derive the intervals of also possible values around the crisp evaluations provided by less certain or less experienced evaluators.
- it is designed for repeated measurement
- for low values of k it provides a graphical interface to present the results to the evaluators
- the k-dimensional representation of the final evaluation of the object does not require aggregation across criteria
- ordering of the alternatives can be obtained applying a suitable distance (interval-valued, if needed) of the k-dimensional representations of the evaluations of the alternatives and the k-dimensional representation of an ideal or desirable alternative (its evaluation). The distances from the least desirable alternative (its evaluation) can also be reflected 'TOPSIS-style'. These distances can reflect also the weights of criteria, or even be based on OWA operators as proposed in the linguistic OWA-TOPSIS [60].

As such GSD-based multiple-criteria evaluation seems to be particularly promising in areas where uncertainty of the evaluations is to be expected and needs to be reflected somehow. The design of the method and the application of the kappa function in combination with a simple assessment of (un)certainty or relevance of the provided evaluation is particularly suitable for those evaluation problems where laymen (in terms of risk/uncertainty representations) are evaluating, and where either less tangible or less usual criteria are being used, or where the alternatives are complex, abstract or novel in some way. The area of design management and design evaluation comes to mind as a first representative [53], [61]. But the applications are much wider and include social sciences and business in general, the evaluation of alternatives with emotional value for the evaluators, the assessment of risk, etc.

What seems to be an even stronger argument speaking in favor of the application of the GSD framework in the multiple-criteria evaluation setting is its capability of serving as a multi-expert evaluation analysis tool. In the multi-expert evaluation problem, we can assume the same that we did in the multiple-criteria evaluation setting, plus the fact that the evaluations are being provided by more evaluators and all their views/evaluations need to be reflected in the final decision to some extent. If we assume k criteria are used and m experts are involved in the evaluation task, then the evaluation of a single alternative can be represented by m k-dimensional objects in the k-dimensional Cartesian evaluation space. Apart from the desirable properties of GSD listed before, we can now consider also:

- the ability to see potential clusters of experts with similar evaluations - their number, distance etc. Obviously clustering techniques can be applied directly to the k-dimensional evaluations to define the clusters, if needed. This can bring understanding concerning the composition of the set of evaluators in terms of their priorities, mutual agreement or even the number of potential points of view on the evaluation

- the overall evaluation of the alternative does not need to be represented as a single k-dimensional object in the evaluation space calculated as an average of all the expert evaluations (applying some aggregation operator) but instead can be represented by:
  - centroids of clusters of experts with similar background/opinion or simply similar evaluations of the alternative (information summarization such that different groups of experts and their views/evaluations remain visible)
  - by the set of m $k-$dimensional evaluations (no information reduction)
  - by a 'union' of the m $k-$dimensional evaluations constructed e.g. as a minimum k-dimensional evaluation such that all the other evaluations are its subsets in the k-dimensional space (maximally careful but potentially very uncertain summary)
  - by an 'intersection' of the m $k-$dimensional evaluations (if a nonempty intersection exists) - this would represent the 'common ground' or 'full agreement' of the experts in terms of their evaluations
  - by an evaluation of a specified shape (in the k-dimensional space) that is the closest to all the other evaluations (ideal compromise)
  - etc.
- the benefit from the possibility of finding consensus of expert evaluations [61] (intersections of the evaluations within a specified (sub) group of evaluators) and analyzing the compatibility of expert assessments by investigating the intersections of the k-dimensional evaluations either overall or dimension by dimension, or the distances of the expert evaluations from each other, from the centroids of clusters (if available), etc. Stoklasová et al. [61] define various types of consensus of expert evaluations that can be applied in the multiple-criteria multi-expert evaluation setting using the GSD evaluation method.

Overall the method allows for various definitions of the overall evaluation of the alternative based on m expert evaluations including such that lose very little information, it allows for the identification of (non) existence of the consensus of expert evaluations (overall and in terms of specific criteria), and for the identification of various types of consensus proposed in [61]. It can be used not only for the determination of the final group evaluation, but also for the analysis of the group of evaluators based on their evaluations. From the above mentioned points it seems that the GSD evaluation applied in the multi-expert setting can prove to be a useful tool for the evaluation and also for the understanding of the evaluation process.

V. Conclusions

Given the above mentioned analysis of the main features and potential benefits of the use of the GSD in multiple-criteria and multi-expert evaluation, the tool seems to be a reasonable candidate for future research concerning its applicability in this domain. We have managed to identify and stress some possible benefits of the use of this tool including the ability to assess less tangible aspects, the ability to model uncertainty, a convenient way of the presentation of results, simplicity of obtaining inputs etc. We have also analyzed the requirements of the method and there do not seem to be any major drawbacks preventing the applicability of the GSD-based tools in multiple-criteria and multi-expert evaluation problems. We have outlined a possible way to apply the main ideas of GSD in this context. More detailed description of the applicability of the method and practical application studies will be the subject of future research.

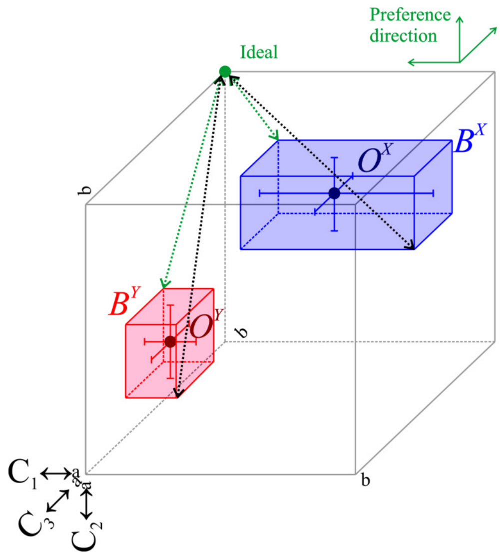

Fig. 3. An example of the output of the generalized semantic differential method [54] for two objects/concepts X and Y in the multiple-criteria evaluation setting. Three criteria C_1, C_2 and C_3 are assumed. The two evaluated objects X and Y are represented by O^X and O^Y and by the "boxes of uncertainty" B^X and B^Y around the O^x and O^Y respectively. The preference direction for all three criteria is shown in green and an ideal evaluation is defined based on this. Green dashed arrows denote the closest Euclidean distances from the ideal to the boxes of uncertainty, black ones the largest distance from the ideal to the points in the boxes of uncertainty.

References

[1] G. Mitra, H. J. Greenberg, F. A. Lootsma, M. J. Rickaert, and H.-J. Zimmermann, Eds., *Mathematical Models for Decision Support*. Berlin Heidelberg New York London Paris Tokyo: Springer-Verlag, 1988, vol. F48. ISBN 978-0387500843

[2] J. Figueira, S. Greco, and M. Ehrgott, Eds., *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer, 2005. ISBN 0387230815

[3] M. Doumpos, J. R. Figueira, S. Greco, and C. Zopounidis, Eds., *New Perspectives in Multiple Criteria Decision Making: Innovative Applications and Case Studies*. Springer Nature Switzerland AG, 2019.

[4] G. Silahtaroglu, H. Dinçer, and S. Yüksel, *Data Science and Multiple Criteria Decision Making Approaches in Finance: Applications and Methods*. Cham: Springer Nature Switzerland AG, 2021.

[5] G.-H. Tzeng and K.-Y. Shen, *New Concepts and Trends of Hybrid Multiple Criteria Decision Making*. New York: CRC Press, 2017. ISBN 978-1-4987-7708-7

[6] C. Zopounidis and M. Doumpos, Eds., *Multiple criteria decision making - Applications in Management and Engineering*. Springer International Publishing AG Switzeland, 2017.

[7] M. Metfessel, "A proposal for quantitative reporting of comparative judgements," *Journal of Psychology*, vol. 24, no. 2, pp. 229–235, 1947.

[8] B. F. Hobbs, "A Comparison of Weighting Methods in Power Plant Siting," *Decision Sciences*, vol. 11, no. 4, pp. 725–737, 1980. doi: 10.1111/j.1540-5915.1980.tb01173.x

[9] T. L. Saaty, "How to Make a Decision: The Analytic Hierarchy Process," *European Journal of Operational Research*, vol. 48, no. 1, pp. 9–26, 1990.

[10] L. A. Zadeh, "Similarity relations and fuzzy orderings," *Information sciences*, vol. 3, no. 2, pp. 177–200, 1971.

[11] L. A. Zadeh and J. Kacprzyk, Eds., *Computing with words in information/intelligent systems 2: Applications*. Berlin Heidelberg GmbH: Springer-Verlag, 1999. ISBN 9783790824612

[12] J. Mingers and L. White, "A review of the recent contribution of systems thinking to operational research and management science," *European Journal of Operational Research*, vol. 207, no. 3, pp. 1147–1161, 2010. doi: 10.1016/j.ejor.2009.12.019

[13] P. Holeček, J. Talašová, and I. Müller, "Fuzzy Methods of Multiple-Criteria Evaluation and Their Software Implementation," in *Cross-Disciplinary Applications of Artificial Intelligence and Pattern Recognition: Advancing Technologies*, V. K. Mago and N. Bhatia, Eds., Hershey, 2012, pp. 388–411. ISBN 9781613504291

[14] Z. Gong, Y. Lin, and T. Yao, *Uncertain Fuzzy Preference Relations and Their Applications*. Heidelberg New York Dordrecht London: Springer-Verlag, 2013. ISBN 9783642284472

[15] A. Sen, "Behaviour and the Concept of Preference," *Economica*, vol. 40, no. 159, pp. 241–259, 1973. doi: 10.2307/2552796

[16] Z. Xu, "A method based on linguistic aggregation operators for group decision making with linguistic preference relations," *Information Sciences*, vol. 166, no. 1, pp. 19–30, oct 2004. doi: 10.1016/j.ins.2003.10.006

[17] E. Herrera-Viedma, J. L. García-Lapresta, J. Kacprzyk, M. Fedrizzi, H. Nurmi, and S. Zadrozny, Eds., *Consensual processes*. Berlin Heidelberg: Springer, 2011. ISBN 9783642133428

[18] S. Greco, B. Bouchon-Meunier, G. Coletti, M. Fedrizzi, B. Matarazzo, and R. R. Yager, Eds., *Advances in Computational Intelligence: 14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU 2012, Catania, Italy, July 9-13, 2012, Proceedings, Part IV*. Heidelberg New York Dordrecht London: Springer, 2012.

[19] G. Tzeng and J. Huang, *Multiple Attribute Decision Making: Methods and Applications*. Boca Raton, London, New York: CRC Press, Taylor & Francis group, 2011. ISBN 978-1-4398-6157-8

[20] J. Krejčí and J. Stoklasa, "Aggregation in the analytic hierarchy process: Why weighted geometric mean should be used instead of weighted arithmetic mean," *Expert Systems with Applications*, vol. 114, pp. 97–106, 2018. doi: 10.1016/j.eswa.2018.06.060

[21] V. Jandová, J. Krejčí, J. Stoklasa, and M. Fedrizzi, "Computing interval weights for incomplete pairwise-comparison matrices of large dimension - a weak consistency based approach," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 6, pp. 1714–1728, 2017. doi: 10.1109/TFUZZ.2016.2633364

[22] T. L. Saaty, "A scaling method for priorities in hierarchical structures," *Journal of Mathematical Psychology*, vol. 15, no. 3, pp. 234–281, jun 1977. doi: 10.1016/0022-2496(77)90033-5

[23] T. Talášek and J. Stoklasa, "Ordering of fuzzy quantities with respect to a fuzzy benchmark - how the shape of the fuzzy benchmark and the choice of distance / similarity affect the ordering," in *Proceedings of the 36th International Conference on Mathematical Methods in Economics*, L. Váchová and O. Kratochvíl, Eds. Jindrichuv Hradec: MatfyzPress, 2018. ISBN 978-80-7378-372-3 pp. 573–578.

[24] C. Hwang, Y. Lai, and T. Liu, "A new approach for multiple objective decision making," *Computers & Operations Research*, vol. 20, no. 8, pp. 889–899, 1993. doi: 10.1016/0305-0548(93)90109-V

[25] L. Basile and L. D'Apuzzo, "Transitive matrices, strict preference order and ordinal evaluation operators," *Soft Computing*, vol. 10, no. 10, pp. 933–940, 2006. doi: 10.1007/s00500-005-0020-z

[26] R. R. Yager, "An approach to ordinal decision making," *International Journal of Approximate Reasoning*, vol. 12, no. 3-4, pp. 237–261, 1995.

[27] J. Stoklasa, T. Talášek, and P. Luukka, "Fuzzified Likert scales in group multiple-criteria evaluation," in *Soft Computing Applications for Group Decision-making and Consensus Modeling*, M. Collan and J. Kacprzyk, Eds. Springer International Publishing AG, 2018, vol. 357, pp. 165–185. ISBN 978-3-319-60206-6

[28] G. Bojadziev and M. Bojadziev, *Fuzzy logic for business, finance and management*, 2nd ed. World Scientific, 2007.

[29] J. Talašová, *Fuzzy metody vícekriteriálního hodnocení a rozhodování*. Olomouc: Palacký University in Olomouc, 2003. ISBN 8024406144

[30] H.-J. Zimmermann, *Fuzzy Set Theory and Its Applications*. Boston/Dordrecht/London: Kluwer Academic Publishers, 2001. ISBN 0792374355

[31] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Aplications*. New Jersey: Prentice Hall, 1995. ISBN 0131011715

[32] M. Smithson and J. Verkuilen, *Fuzzy set theory: applications in the social sciences*. Thousand Oaks, London, New Delhi: Sage Publications, 2006. ISBN 076192986X

[33] V. Traneva, S. Tranev, and D. Mavrov, "Interval-Valued Intuitionistic Fuzzy Decision-Making Method using Index Matrices and Application in Outsourcing," in *Proceedings of the 16th Conference on Computer Science and Intelligence Systems, ACSIS, Vol. 25*, M. Ganzha, L. Maciaszek, M. Paprzycki, and D. Ślęzak, Eds., 2021. doi: 10.15439/2021F77 pp. 251–254.

[34] T. T. Yaman, "Pythagorean Fuzzy Analytical Network Process (ANP) and Its Application to Warehouse Location Selection Problem," in *Proceedings of the Federated Conference on Computer Science and Information Systems, ACSIS, Vol. 21*, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., 2020. doi: 10.15439/2020F187 pp. 137–140.

[35] F. Herrera and E. Herrera-Viedma, "Linguistic decision analysis: steps for solving decision problems under linguistic information," *Fuzzy Sets and Systems*, vol. 115, no. 1, pp. 67–82, oct 2000. doi: 10.1016/S0165-0114(99)00024-X

[36] J. Stoklasa, *Linguistic models for decision support*. Lappeenranta: Lappeenranta University of Technology, 2014. ISBN 978-952-265-686-2 (PDF)

[37] J. Kacprzyk and S. Zadrozny, "Linguistic Data Summarization: A High Scalability through the Use of Natural Language?" in *Scalable Fuzzy Algorithms for Data Management and Analysis: Methods and Design*, A. Laurent and M.-J. Lesot, Eds. IGI Global, 2010, pp. 214–237.

[38] M. Delgado, J. L. Verdegay, and M. A. Vila, "Linguistic decision-making models," *International Journal of Intelligent Systems*, vol. 7, no. 5, pp. 479–492, 1992. doi: 10.1002/int.4550070507

[39] M. Collan and J. Kacprzyk, Eds., *Soft Computing Applications for Group Decision-making and Consensus Modeling*. Cham: Springer international publishing AG, 2018. ISBN 978-3-319-60206-6

[40] V. Sukač, J. Talašová, and J. Stoklasa, "'Soft' consensus in decision-making model using partial goals method," in *Proceedings of the 34th International Conference on Mathematical Methods in Economics*. Liberec: Technical University of Liberec, 2016. ISBN 978-80-7494-296-9 pp. 791–796.

[41] M. Fedrizzi, M. Fedrizzi, R. A. M. Pereira, and M. Brunelli, "Consensual dynamics in group decision making with triangular fuzzy numbers," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2008. doi: 10.1109/HICSS.2008.100. ISBN 0769530753. ISSN 15301605 pp. 1–9.

[42] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases," *Science*, vol. 185, no. 4157, pp. 1124–1131, 1974.

[43] D. Kahneman, "The Semantic Differential and the Structure of Inferences Among Attributes," *The American journal of psychology*, vol. 76, no. 4, pp. 554–567, 1963.

[44] A. Tversky and D. Kahneman, "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, vol. 5, no. 4, pp. 297–323, 1992. doi: Doi 10.1007/Bf00122574

[45] ——, "The framing of decisions and the psychology of choice." pp. 453–458, 1981.

[46] R. P. Hämäläinen, J. Luoma, and E. Saarinen, "On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems," *European Journal of Operational Research*, vol. 228, no. 3, pp. 623–634, aug 2013. doi: 10.1016/j.ejor.2013.02.001

[47] J. Stoklasa, T. Talášek, and J. Stoklasová, "Executive summaries of uncertain values close to the gain/loss threshold - linguistic modelling perspective," *Expert Systems with Applications*, vol. 145, p. 113108, 2020. doi: 10.1016/j.eswa.2019.113108

[48] J. Stoklasa and M. Kozlova, "What is your problem, decision maker? Do we even care anymore?" in *KNOWCON 2020, Knowledge on Economics and Management, Conference Proceedings*, P. Slavíčková

and J. Stoklasa, Eds. Olomouc: Palacký University Olomouc, 2020. doi: 10.5507/ff.20.24457987. ISBN 9788024457987 pp. 208–214.

[49] A. Morreale, J. Stoklasa, and T. Talášek, "Fuzzy Grouping Variables in Economic Analysis. A Pilot Study of a Verification of a Normative Model for R&D Alliances," *Fuzzy Economic Review*, vol. 21, no. 2, pp. 19–46, 2016.

[50] Y. Mitev and L. Kirilov, "Group Decision Support for e-Mail Service Optimization through Information Technology Infrastructure Library Framework," in *Proceedings of the 16th Conference on Computer Science and Intelligence Systems, ACSIS, Vol. 25*, M. Ganzha, L. Maciaszek, M. Paprzycki, and D. Ślęzak, Eds., 2021. doi: 10.15439/2021F93 pp. 227–230.

[51] R. Calvo, S. D'Mello, J. Gratch, and A. Kappas, Eds., *The Oxford Handbook of Affective Computing*. Oxford University Press, jan 2015. ISBN 9780199942237

[52] C. E. Osgood, G. J. Suci, and P. H. Tannenbaum, *The Measurement of Meaning*. Chicago: University of Illinois Press, 1957. ISBN 0252745396

[53] J. Stoklasa, T. Talášek, and J. Stoklasová, "Reflecting emotional aspects and uncertainty in multi-expert evaluation: one step closer to a soft design-alternative evaluation methodology," in *Advances in Systematic Creativity: Creating and Managing Innovations*, L. Chechurin and M. Collan, Eds. Palgrave Macmillan, 2019, pp. 299–322. ISBN 978-3-319-78074-0

[54] ——, "Semantic differential for the twenty-first century: scale relevance and uncertainty entering the semantic space," *Quality & Quantity*, vol. 53, no. January 2019, pp. 435–448, may 2019. doi: 10.1007/s11135-018-0762-1

[55] C. E. Osgood, "Semantic Differential Technique in the Comparative Study of Cultures," *American Anthropologist*, vol. 66, no. 3, pp. 171–200, 1964. doi: 10.1515/9783110215687.109

[56] O. Friborg, M. Martinussen, and J. H. Rosenvinge, "Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience," *Personality and Individual Differences*, vol. 40, no. 5, pp. 873–884, 2006. doi: 10.1016/j.paid.2005.08.015

[57] N. Kervyn, S. T. Fiske, and V. Y. Yzerbyt, "Integrating the stereotype content model (warmth and competence) and the Osgood semantic differential (evaluation, potency, and activity)," *European Journal of Social Psychology*, vol. 43, no. 7, pp. 673–681, 2013. doi: 10.1002/ejsp.1978

[58] D. R. Heise, "Some methodological issues in semantic differential research." *Psychological Bulletin*, vol. 72, no. 6, pp. 406–422, 1969. doi: 10.1037/h0028448

[59] J. D. Dombi and A. Kertész, "Advanced Scheduling Techniques with the Pliant System for High-Level Grid Brokering," in *Agents and Artificial Intelligence Second International Conference, ICAART 2010*, J. Filipe, A. Fred, and B. Sharp, Eds. Heidelberg New York Dordrecht London: Springer, 2011, pp. 173–185. ISBN 9783642198892

[60] P. Luukka and J. Stoklasa, "Similarity based topsis with linguistic-quantifier based aggregation using owa," in *Recent Advances in Business Analytics. Selected papers of the 2021 KNOWCON-NSAIS workshop on Business Analytics*, ser. Annals of Computer Science and Information Systems, J. Stoklasa, P. Luukka, and M. Ganzha, Eds., vol. 29. PTI, 2021. doi: 10.15439/2021B6 p. 45–51. [Online]. Available: http://dx.doi.org/10.15439/2021B6

[61] J. Stoklasová, T. Talášek, and J. Stoklasa, "Attitude-based multi-expert evaluation of design," in *Intelligent Systems and Applications in Business and Finance*, P. Luukka and J. Stoklasa, Eds. Springer, 2022, p. (in press).