QUANTITATIVE METHODS IN ECONOMICS
Vol. XV, No. 2, 2014, pp. 284 – 295

TRANSFORMATIONS OF LINGUISTIC AND NUMERICAL DATA IN THE COMPUTERISED DECISION SUPPORT SYSTEM

Jarosław Becker
The Department of Technology
The Jacob of Paradyż University of Applied Sciences in Gorzów Wielkopolski
e-mail: jbecker@pwsz.pl
Ryszard Budziński
Faculty of Economics and Management, University of Szczecin
e-mail: ryszard.budzinski@wneiz.pl

Abstract: The article discussed the issue of the bi-directional data transformation (from the linguistic form into numerical, and vice versa) in the computerised decision support system (DSS). The system uses multi-methodical research approach, which is to provide a simple and useful form of functionality of different, complementary decision support methods. Data transformations are based on the use of fuzzy set logic and specially developed for thus purpose scoring and linguistic scales of the ordinal nature.

Keywords: linguistic and numerical data transformations, fuzzy quantifiers, computerised decision support system.

INTRODUCTION

In business practice very common are situations, in which information expressed orally are necessary to make a decision. This type of expression form is cognitive in nature and closer to the human perception of reality. By nature, we perceive and describe objects and phenomena in the imprecise and blurred way. Only the need to make precise calculations, for example, of the engineering nature, forces the use of right tools and measurement methods and expressing some properties using precise numerical values.

Literature [Bouyssou, Roy 1993], [Greco et al. 2001, 2002], [Słowiński 2007] contains a variety of procedures and methods of multiple criteria decision making (MCDM). According to Greco et al. (2001) they can be divided into methods based on the functional model (American school) and relational model (European school). The vast majority of these methods depends on the input data expressed numerically. The remaining group, constituting the complement in this context, are the research methods created on the basis of statistics, artificial intelligence and psychology, in which the numerical parameters characterising the research subject are not specified (phenomenon, object). They are called the non-parametric methods, often there are no assumptions in them as to the completeness or precision of data. This group, for example, includes the symbolic methods of data classification [Gatnar 1998] and most of the methods based on the theory of rough sets, applied to the analysis of data consistency, their grouping and induction of decision-making rules [Pawlak 1982].

Integration of many complementary methods of decision-making in the information system requires, first of all, the development of such a model of data organization which will be more adjusted to the theory of decision-making. Secondly, the integration requires arming of the decision-making analysis process on its each step with computer algorithms of transformation of various data forms in such a way that in the context of the problem there is used one common set of input data (numeric, linguistic or mixed).

The article focused on the issue of bi-directional transformation of linguistic and numerical data, which was used in the computerised decision support system – DSS (version DSS 2.0 – author Budziński R., Becker J., 2008-2014). This system distinguishes the multi-methodical research method, consisting of sharing the simple and useful form of algorithms of different, complementary decision support methods. It covers the issues of selection, ordering and grouping of objects (decision variants) from the point of view of the determined set of criteria and preferences and the possibility to take in to account the set of restrictive conditions. Apart from this, it enables the econometric valuation of objects and induction of decision rules. Data transformations were based on the use of the fuzzy set logic and specially developed for this purpose profiles of ordinal and linguistic scales. The broader context for this data conversion in the computer system is the integration of knowledge sources – measurement data, expert opinions, unified structures of mathematical models and collections of methods – at an important rime of the information and decision process, that is the decision game, which purpose is the selection of the best solutions from the available ones.

FUNCTIONAL SCOPE OF THE COMPUTERISED DECISION SUPPORT SYSTEM

The functional scope of supporting the decisions was determined as the solving of decisive tasks connected with multi-criteria selection, grouping (sorting) and organising (ranking) of any decision variants, understood as objects of the analysis representing the given category of events or things. These objects must have a uniform information structure. The additional functionality of the system is the analysis and the evaluation *ex post* of the obtained results of the decision-making process. It should be noted that the studies carried out in the system can have the formal nature (official), taking on the form of the legally sanctioned procedure (e.g. public tender, where the offers are evaluated) or less official, cognitive, where the decision maker is repeatedly supported through simulations (e.g. evaluation of employees, products, services, variants of planning, etc.). The fact that the theory of decisions creates methodological foundations for the analysis and generating best solutions is not about the utility of the information system in practice. In fact, the needs of management translate into the essential factors that should be taken into account in the design of system supporting decision-making, namely:

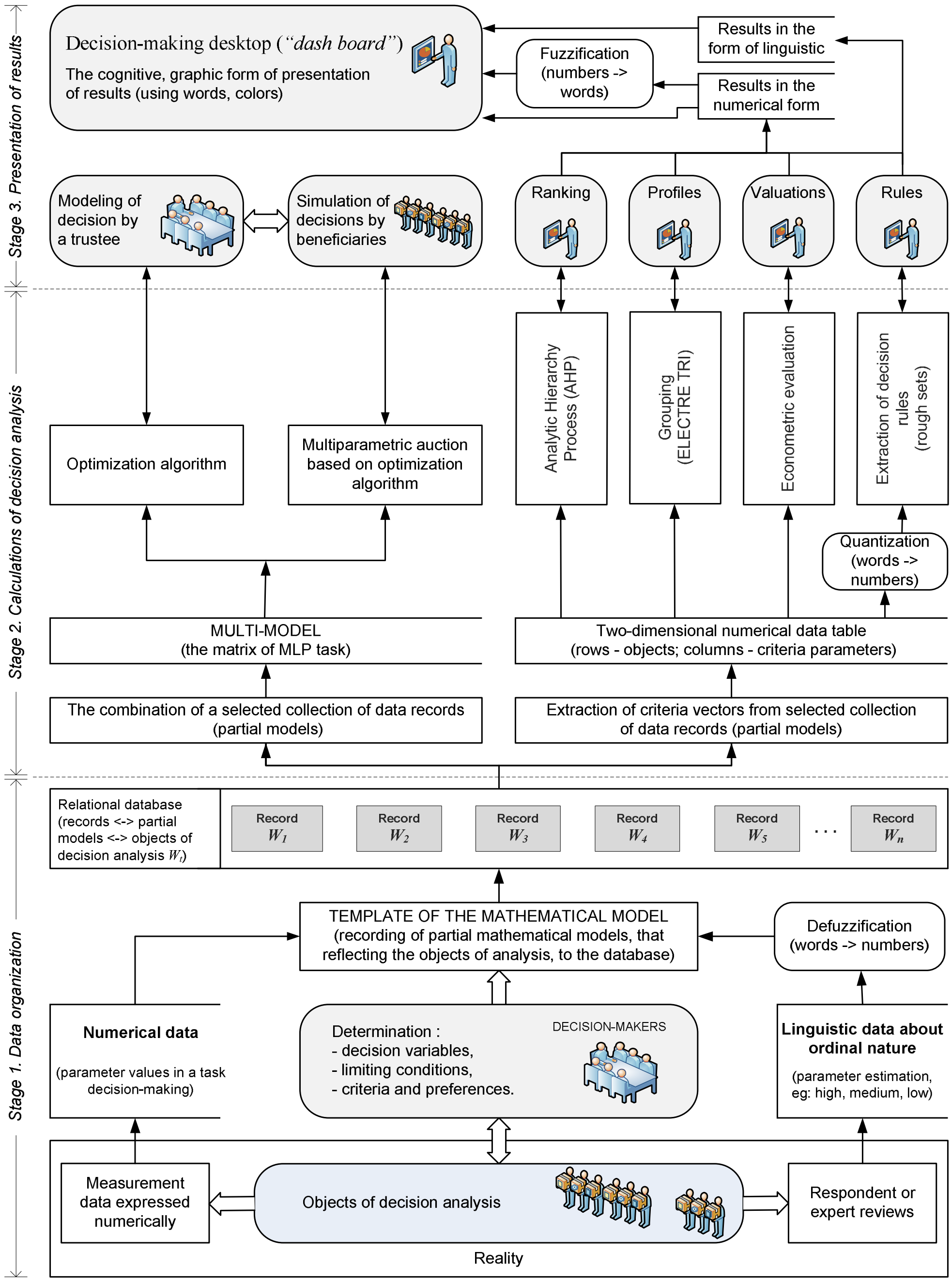
- multi-stage nature of the decision-making process,
- multi-criteria nature, in which the structure of criteria is simple (criteria vector) or complex (hierarchical or network dependencies),
- number of decision-makers and experts,
- scale of the decision problem (few or mass problems),
- flexibility of decision variants (customising the parameter values),
- linguistics of data (statements of experts or respondents).

The complexity of the description of the decisive situation causes that it is difficult to emerge the method that would be universal, to which we could attribute the possibility to obtain the best solution of many different decision-making problems.

The discussed system of supporting decision-making is a hybrid solution, which using the engineering techniques of the computer processing of data connects and shares in a simple useful form algorithms of various supplementary and implementing the paradigm of the methods supporting the decisions. The research procedure included in it is performed in three stages, it includes: (1) organization of data, (2) calculations of the decision analysis and (3) presentation of results (Figure 1). The intention of the proposed scheme of thought comes from the understanding of the support of decisions as a process, in which based on the fact base (data) we analyse and conclude, and then we make decisions. This takes into account the knowledge of users and most of all of experts, who analyse facts, express their opinions using the ordinal scale of linguistic assessments and use the mapping methods proposed in the system.

Organizing data (Figure 1) as the base of integration of methods there was accepted the coherent and flexible information structure of the system, which was subordinated to the construction of MLP models (Multi-criteria Linear Programming). It allows you to define the template for the decision-making task (standard mathematical model, Figure 1). This construction takes into account the requirements of the decision maker, which relate to the potentially analysed set of objects and they are expressed through: decision variables, limiting conditions, one- or two-level structure of criteria of assessment and the corresponding preferences [Becker 2008].

Figure 1. General architecture of the computerized decision support system (DSS 2.0)



Source: own work

According to the template to the system there are introduced data of objects (decision variants: W_1, W_2, …, W_n). Technical and economic parameters of each variant can be expressed in the form of numerical values and linguistic assessments (fuzzy values) from the ordinal scale defined by experts or respondents. For the optimization calculations all linguistic forms of data must get transformed into numerical values. The basis for the conversion of verbal expressions into numerical (defuzzification) and vice versa (fuzzification) is the methodology of the construction of linguistic quantifiers based on the theory of fuzzy sets. After the introduction and confirmation of data, each variant becomes the record (writing) in the relational database and at the same time is the autonomous, partial mathematical model. The object takes the form of the formalised task of the linear programming, which after obtaining the positive optimization result (where it is not the contrary system) is saved in the database with the admission status to the stage of decision analysis calculations.

The second stage (Figure 1) includes the issues of combining data records – partial mathematical models identical to objects of the decision-making analysis – to the form of a multi-model (MLP task matrix) for the needs of the multi-criteria optimization and transformation to the simple, tabular structure of data required on other inputs of the multi-methodical analysis. Integration of methods in the system of supporting decisions consists of the use of their functionality on a common set of data (objects) within a coherent, logical and comprehensive information-decisive process consisting of:

A. *optimization of decisions* – considered from the point of view of interests of the trustee's resources and from the perspective of beneficiaries competing for the resources,
B. *multi-criteria analysis*, in which there were used the approaches: connected with the achievements of the American school (AHP method [Saaty 1980]), European (ELECTRE TRI [Roy 1991]) and Polish school (Rough Set Theory – [Pawlak 1982]),
C. *identification* in terms of quantitative methods of the econometric analysis.

The third stage (Figure 1) includes the presentation of detailed results for each method separately and together, in the form of the decision-making desktop *("dash board")*, within which the applied methods (points B and C) function on the basis of a consultation of experts diagnosing the state of the tested objects. The desktop integrates the results of methods supporting decisions in the utility aspect. It is an interactive system enabling the multi-dimensional (multi-methodical) diagnostics of the selected object W_t (or a new one W_{n+1}) against the results of the whole set (W_1, W_2, …, W_n). It has the cognitive, graphic form of presentation of results of the applied methods. It is a kind of machine graphics, which consolidates the graphic visualization with cognitive processes taking place in the man's mind at the moment of making the decision. The structure of the desktop is based on the premise that knowledge about the object (its rating) expressed by shape and colour is absorbed faster than information in the form of numbers and text.

THE USE OF LINGUISTIC QUANTIFIERS FOR THE CONSTRUCTION OF ORDINANCE SCALE PROFILES

In the computer decision support system there are distinguished three areas of the use of linguistic quantifiers. The first one involves the transformation of linguistic data (e.g. assessment of experts) to the numerical form (defuzzification) required in the MLP mathematical model. It may relate to all values or only some parameters (defined in the template of the decision task, Figure 1), which characterise the W_t objects considered in the decision task. In the second area there is the quantification of data (also referred to as discretisation, Figure1), which is carried out for the purpose of induction of decision rules. The basic quantification stage, connected with the division of the scope of attribute values into separate sections, can be implemented in an automated manner using the selected scoring and linguistic scale or arbitrarily determined by the user. The third area of the use of linguistic quantifiers is the decision display (Figure 1). Data transformations are used in order to unify, consolidate and cognitively visualise the analysis results, obtained with different methods.

The ideal, to which we aspire, are the lossless transformations of numerical data to the linguistic forms, and vice versa. Tools, which relatively well allow for this type of conversions, are provided by the fuzzy logic. It is based on the term of *fuzzy sets*, meaning those, which do not have strictly defined boundaries. In 1965, Zadeh provided an idea and the first concept of a theory, enabling the fuzzy description of real systems. The fuzzy set is an object including the elements of some area of considerations, wherein each of these elements can fully belong to the fuzzy set, do not belong to it at all or belong to it to some degree [Łachwa 2001]. The fuzzy set A in the space (area of considerations) $X = \{x\}$, what can be written as $A \subset X$, is a set of pairs

$$A = \{(f_A(x), x)\}, \quad \forall x \in X, \tag{1}$$

where f_A: $X \rightarrow [0, 1]$ is the membership function, which assigns each element of the X space with the grade of membership to the given fuzzy set: from membership $f_A(x) = 0$ through partial membership $0 < f_A(x) < 1$ to complete membership $f_A(x) = 1$ [Kacprzyk 1986].

In fuzzy sets, the transition from membership to non-membership is gradual, and not abrupt, as in the conventional set. The concept of a fuzzy set is used for the formal recognition and quantitative expression of blurry, imprecise, ambiguous terms. They are commonly used for the qualitative assessment of physical quantities, conditions of objects and systems, and their comparison [Piegat 1999].

The concept of the scoring and linguistic scale profile ($scale^{(\tau)}$) in the computerised decision support system means the user-determined configuration of the adjustable elements of the ordinal scale, i.e.:

- *number of degrees* $\tau = 2, 3, \ldots, 11$ – the system distinguishes 9 variants of the span of $scale^{(\tau)}$ (they were given Latin names: 'duo', 'tria', 'quatuor', ‘quinque’, etc., in which the next degrees were given the absolute, non-negative ordinance values $\alpha = 0, 1, 2, \ldots, \tau\text{-}1$, always starting from zero,

- *linguistic values (names of degrees)* $a^{(\alpha)}$ – for example, for $scale^{(\tau=3)}$ these may include: $a^{(\alpha=0)}$ = *‘low’*, $a^{(\alpha=1)}$ = *‘average’*, $a^{(\alpha=2)}$ = *‘high’*,
- *type of characteristics of the linguistic quantifier* – this is a non-linear dependency ($y = ax^2 + bx$) or linear ($y = x$) applied to generate, for any span of $scale^{(\tau)}$, triangular or pentagonal membership functions for individual linguistic values $a^{(\alpha)}$. [Becker 2014].

Linguistic quantifier consists of the membership functions, which number corresponds to the number of degrees τ on the given scale. These functions are created based on linear or non-linear function transformations. For each scale there can be determined many different linguistic quantifiers. In the computer system there are prepared five basic variants of the linguistic quantifier with: a) proportional, b1) strongly growing, b2) moderately growing, c1) strongly decreasing and c2) moderately decreasing distances between linguistic values $a^{(\alpha)}$. Disproportionate versions (b1, b2, c1, c2) can consist of triangles or pentagons, what in total gives nine proposals. For advanced users there is predicted the possibility of adjusting the shape of the characteristics of the linguistic quantifier according to the relationship

$$f_\eta(x) = y = (\eta\text{-}1)x^2 + (2\text{-}\eta)x, \tag{2}$$

in which the η parameter adopts the values from the range of $\langle 0; 2\rangle$, and $x = \alpha/(\tau\text{-}1)$. If $\eta = 1$, the characteristics (2) is linear, distanced obtained on its basis between $a^{(\alpha)}$ are identical, and membership functions for each $a^{(\alpha)}$ have the form of equilateral triangles (except for extreme values $a^{(0)}$ and $a^{(\tau\text{-}1)}$, for which the half-figures are always taken into account). In other cases (when $\eta \neq 1$) the inscription (2) determines non-linear relationships, and membership functions obtain the selected shape, of a triangle or pentagon. When $\eta \in \langle 0; 1)$, the system generates quantifiers of decreasing distances between successive $a^{(\alpha)}$ ($\alpha = 0, 1, 2, \ldots, \tau\text{-}1$), while for $\eta \in (0; 2\rangle$ proportions of these distances move in the opposite direction.

In order to simplify the notation of the function determining the degree of variable membership $x' \in \langle 0; 1\rangle$ to linguistic values $a^{(\alpha)}$, determined for particular degrees $\alpha = 0, 1, 2, \ldots, \tau\text{-}1$, the functional relationship has been transformed (2) to the following form

$$f_\eta(\alpha) = \frac{(\eta-1)\alpha^2}{(\tau-1)^2} + \frac{(2-\eta)\alpha}{\tau-1}. \tag{3}$$

Triangular membership functions are constructed for any span of $scale^{(\tau)}$ – assuming that the variable $x' \in \langle 0; 1\rangle$ and represents the numerical value subject to conversion into the linguistic form – can be expressed in the form of the following entries (Figure 2, an example for the five-point scale):

- for the first linguistic value $a^{(\alpha)}$ ($\alpha = 0$)

$$f_{a^{(0)}}(x') = \begin{cases} \frac{f_\eta(1)-x'}{f_\eta(1)} & for\ f_\eta(1) \geq x' \geq 0 \\ 0 & for\ other\ x' \end{cases}, \tag{4}$$

- when τ > 2, then for every $a^{(\alpha)}$ satisfying the condition 0 < α < τ-1 particular membership functions can be generalised to the form of

$$f_{a^{(\alpha)}}(x') = \begin{cases} \frac{x'-f_\eta(\alpha-1)}{f_\eta(\alpha)-f_\eta(\alpha-1)} & for\ f_\eta(\alpha-1) \leq x' < f_\eta(\alpha) \\ \frac{f_\eta(\alpha+1)-x'}{f_\eta(\alpha+1)-f_\eta(\alpha)} & for\ f_\eta(\alpha) \leq x' \leq f_\eta(\alpha+1) \\ 0 & for\ other\ \ x' \end{cases}, \tag{5}$$

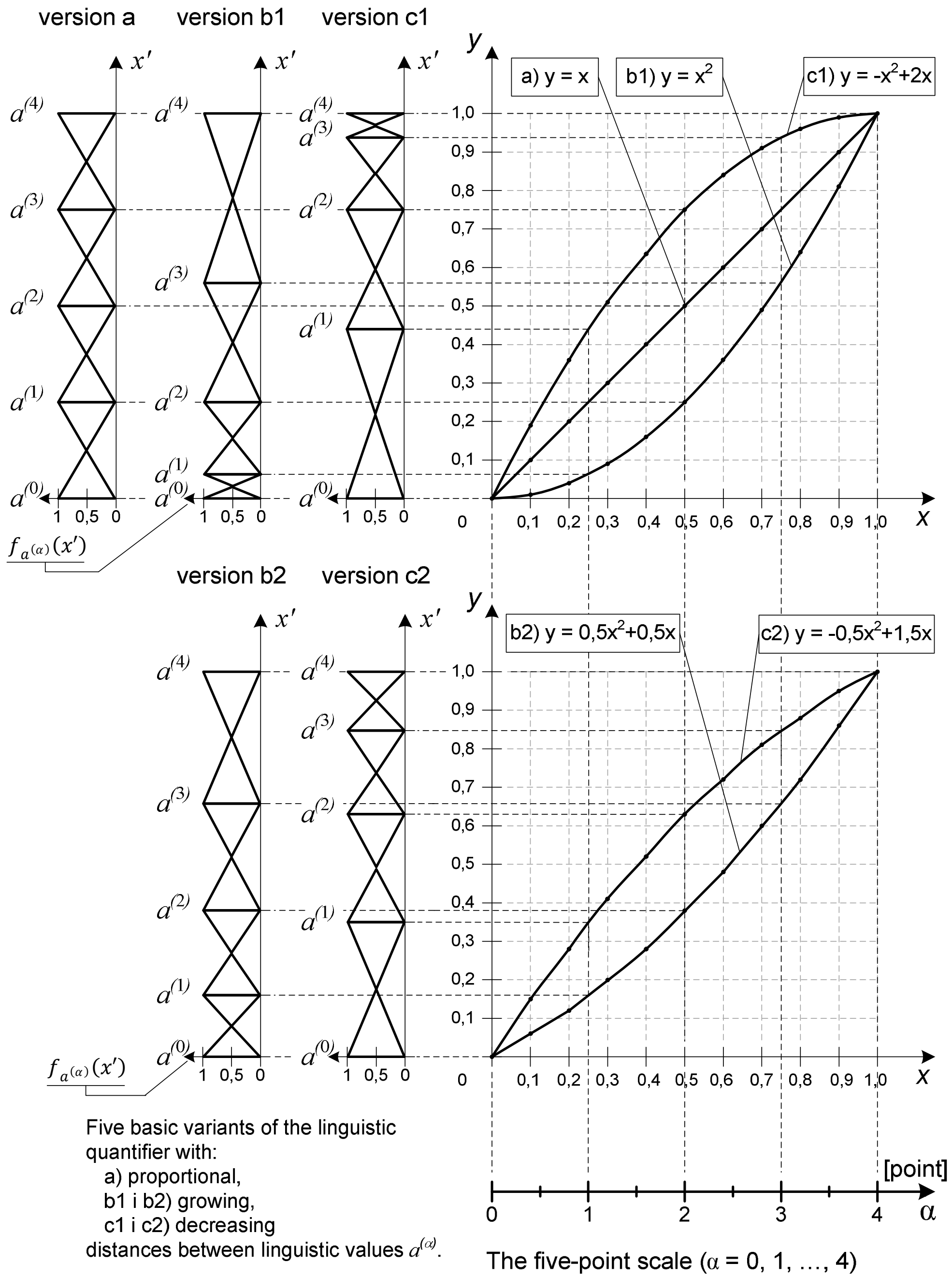
- for the last linguistic value $a^{(\alpha)}$ (α = τ-1)

$$f_{a^{(\tau-1)}}(x') = \begin{cases} \frac{x'-f_\eta(\tau-2)}{1-f_\eta(\tau-2)} & for\ \ f_\eta(\tau-2) \leq x' \leq 1 \\ 0 & for\ other\ \ x' \end{cases}. \tag{6}$$

In a similar manner are constructed the linguistic quantifiers equipped with the membership functions shaped as a pentagon. Pentagonal functions in relation to triangular ones are more approximate to the shape of the non-linear characteristics (2). This is due to the fact that apart from the values $x'_\alpha = f_\eta(\alpha)$ calculated for each α = 0, 1, 2, …, τ-1, for which $f_{a^{(\alpha)}}(x'_\alpha) = 1$ and reaches the extreme, with the same rule there were also determined the intersecting points $x'_{\alpha+} = f_\eta(\alpha + 0{,}5)$ i $x'_{\alpha-} = f_\eta(\alpha - 0{,}5)$ of the adjacent membership functions, where $f_{a^{(\alpha)}}(x'_{\alpha-}) = 0{,}5$ i $f_{a^{(\alpha)}}(x'_{\alpha-}) = 0{,}5$ [Becker 2014].

According to Łachwa (2001) the issue of assigning the parameters describing the specific objects with the right membership degrees to linguistic expressions of the ordinal nature, it should be stated that doing this in a good way is difficult. This procedure is usually of the subjective nature and depends on the situational context. Clarifying this issue, membership degrees which are individual and depend on the circumstances indicate a kind of trend, which reflects on the set of studied objects from the given area of considerations some arrangement, created by association with the set of specific features. To determine the membership degrees there is used, for example, the questionnaire method common in statistics. The membership value is calculated as the relation of the number of affirmative answers to the number of all answers provided by responders. Another, popular method is determination of membership degrees by the expert. However, the expert often determines only the general shape of the membership function, and the accurate parameter values are selected experimentally.

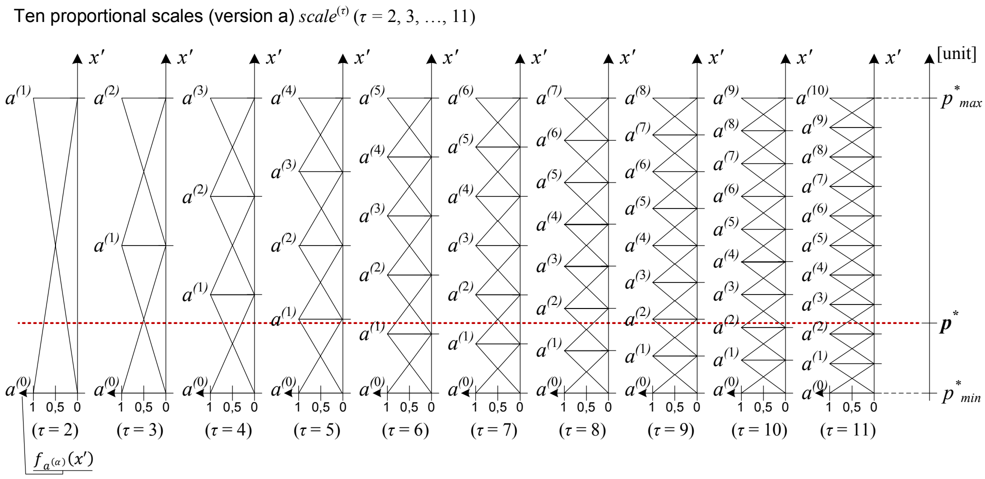
Figure 2. Constructing linguistic quantifiers based on triangular membership functions (example for the five-point scale)



Source: [Becker 2014]

In the decision support system there was proposed a tool for creating individual profiles of scoring and linguistic scales, which act as ready-to-use models of the linguistic quantifiers with the selected membership functions. These may include the equilateral triangles (type a – proportional scales, Figure 3), irregular triangles or pentagons (variants of the b and c type) of varying proportions on the ordinal scale, what is understood as different distances between degrees and focal points (the point of intersection of two functions), additionally they can have the growing or decreasing trend. A multitude of parameters configuring the profile of scoring and linguistic scale allows to define in the given decision task (*z*) the individual, required for each p^* parameter in the template of the mathematical model, linguistic quantifier.

Figure 3. The use of the linguistic quantifiers in the DSS system (projection of values of the $p*$ parameter on the axis of the proportional scales)



Source: [Becker 2014]

The choice of the quantifier characteristics can be the result of the expert's suggestion, discussion of a group of several people (e.g. decision-makers and experts) or the survey, which van be performed in the phase of obtaining data (e.g. in the form of additional questions in the offer proposals submitted by beneficiaries representing the objects).

After determining the character of the membership function for the quantifier and for the given scale profile, there should be determined the number of its degrees (τ). In the system there are two kinds of allocation of this size to the given p^* parameter. The group variant, in which τ is determined for all experts assessing p^* and the individual one, where τ is selected by each expert giving opinion on p^* according to his preferred (intuitive, best perceived) structure of the assessment value system. In the computer system there were made available 9 scale models

($scale^{\tau}$), which have from τ = 2 to τ = 11 degrees (Figure 3). It should be noted that the greater number of degrees on the scale may influence the extension of the process of parameter assessment, but instead there is obtained the higher precision of transformation of linguistic notes onto the numerical values, and in a large group of assessing people this may also result in their greater diversity.

The use of linguistic quantifiers in the decision support system is related to the bi-directional conversion of numerical and linguistic data. Defuzzification means the conversion of signals from the qualitative field to the quantitative one. In the system it relates to the parameter assessment (p^*) expressed with imprecise measures $a^{(\alpha)}$ within the determined profile of the scoring and linguistic scale. For the given $a^{(\alpha)}$, according to (3), the value $x'_\alpha = f_\eta(\alpha)$, is calculated, for which $f_{a^{(\alpha)}}(x'_\alpha) = 1$. Then, the value $x'_\alpha \in \langle 0;1 \rangle$ is proportionally converted into the acceptable parameter scope $p^* \in \langle p^*_{min}; p^*_{max} \rangle$ (Figure 3). The conversion process taking place in the opposite direction, where quantitative data (precise) are converted to qualitative is called fuzzification or dissolving. The numerical value of the parameter $p^* \in \langle p^*_{min}; p^*_{max} \rangle$ is transformed proportionally to $x' \in \langle 0;1 \rangle$, then for each scale process (with the selected model $scale^{\tau}$ and quantifier characteristics – type *a, b* or *c*) there are calculated the values of membership functions for every $a^{(\alpha)}$ (α = 0, 1, 2, …, τ-1) according to the entries (4; 5 and 6) for triangular functions (or similarly for the pentagon-shaped functions – more in the paper [Becker 2014]). The highest value $f_{a^{(\alpha)}}(x')$ from the calculated ones determines the linguistic category $a^{(\alpha)}$. It should be noted that the applied data transformation based on the theory of fuzzy sets – in which based on the linear characteristics and different non-linear ones the scopes of membership functions are determined (equal, increasing or decreasing) – is adequate to the process of determining the quantization intervals (discretisation) of the attribute values in the induction studies of decision rules. The generated data scopes can be clarified by the system user.

SUMMARY

The article presented the issues of the linguistic and numerical transformation of data used within the uses of the computer decision support system (DSS 2.0). They most often concern the problems solved with experts, who express their opinions using imprecise terms, for example, assessment: of employees or recruits, grant or loan applications, tenders including the specialised services or devices, etc. The place of experts may be occupied by respondents, e.g., the representative group of students assessing individual departments, directions of teaching.

Linguistic quantifiers, included in the form of fuzzy and ordinal scale profiles, on one hand, are used in order to bring data to the specific form and provide them to the input of the appropriate decision support method. On the other hand, numerical method results are converted to the linguistic form and integrated with others on the decision desktop. The aim of this operation is the synthesis of results obtained with different methods based on the cognitive presentation and interpretation (using words and spectrum of colours).

What is interesting is the use of scale profiles of a different number of degrees describing the conditional attributes and the decision attribute to the search of such quantization, due to which there will be generated the most valuable rules. Generalising the description of attributes (reducing the number of categories), we admittedly influence the structure and consistency of the data set, but we are moving towards the deep knowledge, expecting the rules of a more general content and greater coverage. This rule can be reversed and the shallower knowledge may be sought, that is more precisely formulated rules in the description of reality.

REFERENCES

Becker J. (2008), Architektura informatycznego systemu generowania wielokryterialnych rozwiązań decyzyjnych (cz. 1). Koncepcja budowy modelu WPL oparta na niestandardowych zadaniach decyzyjnych, Seria IBS PAN: Badania Systemowe, Tom 64, Wyd. Instytut Badań Systemowych PAN & Polskie Towarzystwo Badań Operacyjnych i Systemowych, Warszawa.

Becker J. (2014), Integracja źródeł wiedzy w informatycznym systemie wspomagania decyzji (podstawy metodyczne i inżynierskie), maszynopis monografii naukowej po uzyskaniu pozytywnych recenzji, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie.

Bouyssou D., Roy B. (1993), Aide multicritere a la decision: Methodes et cas, „Economica", Paris.

Gatnar E. (1998), Symboliczne metody klasyfikacji danych, PWN, Warszawa.

Greco S., Matarazzo B., Słowiński R. (2001), Rough sets theory for multicriteria decision analysis, European Journal of Operational Research (129).

Greco S., Matarazzo B., Słowiński R. (2002), Axiomatization of utility, outranking and decision-rule preference models for multiple-criteria classification problems under partial inconsistency with the dominance principle, „Control and Cybernetics", vol.: 31/4, p: 1005-1035.

Kacprzyk J. (1986), Zbiory rozmyte w analizie systemowej, PWN, Warszawa, s. 43.

Łachwa A. (2001), Rozmyty świat zbiorów, liczb, relacji, faktów, reguł i decyzji, Akademicka Oficyna Wydawnicza „Elit", Warszawa.

Pawlak Z. (1982), Rough sets, Int. J. Computer and Information Sci., 11, s. 341-356.

Piegat A. (1999), Modelowanie i sterowanie rozmyte, Akademicka Oficyna Wydawnicza EXIT, Warszawa, s. 23 i 28.

Roy B. (1991), The outranking approach and the foundations of Electre methods, Theory and decision, no. 31.

Saaty T. L. (1980), The analytic hierarchy process: Planning, priority setting, resource allocation. McGraw-Hill International Book Co., New York.

Słowiński R. (2007), Podejście regresji porządkowej do wielokryterialnego porządkowania wariantów decyzyjnych, in: Kulczycki P., Hryniewicz O., Kacprzyk J. (eds.), Techniki informacyjne w badaniach systemowych, Wyd. Naukowo-Techniczne, Wyd. Naukowo-Techniczne, Warszawa.

Zadeh L. (1965), Fuzzy sets, "Information and Control" vol. 8, s. 338-353.