SEMANTICS OF THE SCORING SYSTEM IN PROBLEMS OF EXPERT ASSESSMENT AND INVERSE GAME THEORY

Daniel Bosque6a [1], Francesco De Luca[2*]

1Warsaw School of Economics, Poland
2School of Law, University of Leeds, UK
** Corresponding author: deluca.gareem@gmail.com*

Abstract

Data expressed in points (points, ranks) are widely used in qualimetry, pedagogy, medicine, velology, sports, sociology, project management, ecology, and many other subject areas that knowledge engineering deals with [1] . The source of this data is expert judgment. Solving the problems of developing systems of educational, velological, sports and consumer motivation [2] , as well as problems of inverse game theory for multi-agent systems [1] leads to the construction of a set of scores and a method of their aggregation - a system of points [3] . The processing of expert assessments can also be reduced to the construction of such a system. The mathematical theory of measurements offers for it only one of the variants of statistical averaging [4] .

Kewords:
Data threats, points, ranks, qualimetry, sports, assessments.

Introduction

The basis for their application is tradition, and not the criteria arising from the meaning of the problem. This approach not only hides the processing procedure from meaningful control and divorces it from the semantics of the practical problem, but also leads to gross errors (an example of such an error is the assessment of the results of the examination session 5,5,2; the arithmetic mean gives 4, median 5, meaningfully the assessment 2) [3] .

The purpose of the publication: development of a precedent for the transition from the noncritical use of the available mathematical models for processing scores to the semantic construction of a score system with the subsequent transition to a processing procedure corresponding to the meaning of the task.

A deliberate transition to the semantic construction of the score system does not mean a rejection of mathematical methods in favor of arbitrary processing methods. The task of semantic design is to bring the method of processing point estimates in accordance with the meaning of the problem being solved.

The process of semantic construction of the scoring system is considered in relation to the specific task of a comprehensive assessment of the initial state of the client's health when planning velological events . At the same time, the proposed system can be easily adapted to a number of other applications, for example, to problems of pedagogical and technical testing or to the inverse problem of game theory for multi-agent systems.

First of all, we note that the assessment procedure is hierarchical. Each of the functional systems of the body is assessed for a number of functions by a set of points (a system of points) of the following type:

$$Bi = <bi1, bi2, \ldots bij, \ldots, bin>,$$

where i is the index of the corresponding functional system, n is the number of its point estimates (the dimension of the points system)

Such tuples of scores, in turn, form a set

$$Bo = < B\,1, B\,2, .., Bi, \ldots\, Bm >$$

where m is the total number of evaluated functional systems. The task of the resulting assessment is to obtain, on the basis of these data, the resulting score of the client's initial state of health - bo. For now, we will restrict ourselves to discussing only a point estimate. Semantic construction in this case must begin with the definition of the semantics (status) of the lowest score. The semantics of the entire scoring system also depends on it. Let's show this with an example. Let's say that a three-point rating scale is used:

$$bi \in \{\text{"excellent"}, \text{"good"}, \text{"satisfactory"}\},$$

where Є is the membership function of the set { }

Then there are two options for the appointment - the grade "satisfactory" is set: *a)* at the boundary states of function, between an acceptable level of functioning and dysfunction; *b)* with a pronounced tendency. to dysfunction with a sufficient margin of its stability. The processing procedures at all levels of the hierarchy will depend on the accepted option for assigning grades.

Let us first consider option *(a) of* assigning an assessment of "satisfactory", as more realistic in the context of the client's voluntary referral to velological events. The psychological characteristics of the behavior of many subjectively healthy people lead to the fact that they turn to health-improving measures only after they discover in themselves certain manifestations of dysfunctions of one or several functional systems of the body. In this case, it is possible to recommend introducing a fourth grade, for example, “quite satisfactory” as an intermediate between the grades “good” and “satisfactory” for fixing the states corresponding to option *(b)*retaining the name of the grade "satisfactory" for the worst grade for psychological reasons. This will allow, without iatrogenic effects on clients, to use the semantics of the "weak link" as follows

“If min <bi1, bi2,… bij,…, bin> = “ satisfactory ”,

then Bi = "satisfactory", otherwise Bi = res (SP) " ,

where *res (SP)* is the result of the standard processing procedure adopted for option *(a).*

In a milder form, the same semantics apply to the definition of *bo:*

“If (Bi = “ satisfactory ”) & (Bj = “ satisfactory ”) & (i / = j) ,

then bo = "satisfactory", otherwise Bi = res (SP) " ,

where / = "not equal".

In this case, the degree of softening of the semantics of the weak link will be determined by the user's ideas about the systemic connections of functional systems (that is, how many functions must manifest their borderline states in order to give sufficient grounds to recognize the general state of the organism as borderline).

Let's move on to the semantic construction of the standard procedure for processing scores *SP* , which is used in both variants of assigning scores. As shown in [3] and in the introduction, traditional averaging procedures cannot reflect the semantic characteristics of a scoring system. The assignment of assessments by dividing the scale with a large number of points (for example, 100) into a number of intervals does not display them either (in particular, this is how it is recommended to evaluate the results of knowledge testing in the Bologna process). Although, unlike the arithmetic mean, in this case it is not necessary to perform incorrect arithmetic operations with points, such assignment of assessments is based on calculating the total number of impersonal, like numbers, points.

The semantic design of the SP processing procedure is aimed at increasing the information content of the resulting scores. This can be achieved in different ways (as well as the calculation of traditional averages), therefore, we will consider one of the possible approaches to constructing SP , which we will conditionally call the “principle of the quality of points”. An illustrative example of its widespread implementation is the "honors degree criterion":

“If (none of bij Є B i / = “ quite satisfactory ”) & (gi> 0.75 n)
, then B i = "excellent" ("with honors"), otherwise B i = NOT "excellent" ,

where *n* is the total number of grades, *gi* is the number of “excellent” grades of the *i* - th functional system.

This criterion is not based on the total number of impersonal points. It takes into account the quality of the scores, that is, it takes into account in which grades they were received. This semantic relationship can be expressed by the proverb "a fly in the ointment spoils a barrel of honey." It is also justified in relation to the veleological assessment, because in this case a responsible decision is made on the rehabilitation veleological measures. The second characteristic concerns the acceptable number of "good" ratings. It finds its basis in the semantics of the “principle of sustainability”. The level of indifference (instability) is 0.5. In practical calculations that are not associated with a risk to life, the uncertainty of the calculation taken into account by a safety factor of 1.25-1.5. Thus, 0.75 provides an upper level of robustness for excellent scores. If the marks are set by independent experts, then the level of 0.75 gives an "overwhelming majority of votes". The consistent implementation of the above grounds leads to the resulting criterion:

“If (none of the bij Є B i / = “ quite satisfactory ”) & (gi ≥ 0.75 n), then B i =“ excellent ”, otherwise, if (((none of the bij Є B i / = “ completely satisfactory ") & (gi < 0.75 n)) \ / (si ≤ 0.25n) & (gi ≥ si)), then B i =" good " , otherwise B i =" quite satisfactory "",

where *si* is the number of “quite satisfactory” ratings. It is assumed that the ratings “satisfactory” are not applied. The processing of scores is carried out in the Microsoft Excel environment .

In conclusion, we note that the semantic design of the points system does not contradict the mathematical concept of the mean (according to Cauchy) and, in addition to the direct purpose of assessing the state of the body, it makes it possible to use the incentive of the points system to motivate the client to achieve his health improvement [2] . In the same way, taking into account the semantics of scoring systems can increase the efficiency of solving problems of decision-making, classification, ranking, aggregation and inverse game theory [1,4] .

References

1. Russell S., Norstead P. Artificial intelligence. 2nd ed. - M.: Publishing house "Williams", 2006. - 1408s.
2. Ball GA "Motive": clarification of the concept. // Psychological journal, 2004, volume 25, No. 4. S.56-65.
3. Bogdanov N.I. The semantics of the score system in problem communication with a computer // Depon. 03/02/81 at NIIVSH # 31-81 / Ref. on Sat. "Education in higher and secondary schools." - Issue 6. - M .: N II VSh, 1981.S. 35.
4. Orlov A.I. Decision-making theory. - M .: Publishing house "Mart", 2004. - 656 p.
5. Abbo, A. R., Miller, A., Gazit, T., Savir, Y., & Caspi, O. (2020). Technological developments and strategic management for overcoming the COVID-19 challenge within the hospital setting in Israel. Rambam Maimonides Medical Journal, 11(3).
6. Alanzi, T. M. (2021). Gig Health vs eHealth: Future Prospects in Saudi Arabian Health-Care System. Journal of Multidisciplinary Healthcare, 14, 1945.
7. Albers, S., & Rundshagen, V. (2020). European airlines' strategic responses to the COVID-19 pandemic (January-May, 2020). Journal of air transport management, 87, 101863.

8. Amankwah-Amoah, J., Khan, Z., & Osabutey, E. L. (2021). COVID-19 and business renewal: Lessons and insights from the global airline industry. International Business Review, 30(3), 101802.
9. Auth, G., JokischPavel, O., & Durk, C. (2019). Revisiting automated project management in the digital age-a survey of AI approaches. Online Journal of Applied Knowledge Management (OJAKM), 7(1), 27-39.
10. Batra, D. (2020). The impact of the COVID-19 on organizational and information systems agility. Information Systems Management, 37(4), 361-365.
11. Brammer, S., Branicki, L., & Linnenluecke, M. K. (2020). COVID- 19, societalization, and the future of business in society. Academy of Management Perspectives, 34(4), 493-507.
12. Castro, M. P., & Zermeno, M. G. G. (2020). Being an entrepreneur post-COVID-19-resilience in times of crisis: A systematic literature review. Journal of Entrepreneurship in Emerging Economies.
13. Coccia, M. (2022). Preparedness of countries to face COVID-19 pandemic crisis: Strategic positioning and factors supporting effective strategies of prevention of pandemic threats. Environmental Research, 203, 111678.