<<FUZZY>> QUESTIONNAIRES: MORE INFORMATION FOR MORE APLLICATION

GianCarlo Michellone

Abstract

Limits of traditional evaluation scales are discussed. These scales are discreet, discontinuous, pre-defined. Limits of natural language judgements are also underlined: in fact they are characterised by synthesis problems. In order to give a partial solution, a formal elaboration of verbal judgements is proposed, allowed by the use of Fuzzy Logic. Some questionnaires, with fuzzy scales and triangular membership functions of their terms, are applied to evaluate typical cases of schools and companies.

1. Introduction

Evaluation is typical of nature of human beings and therefore it is a continuos -- conscious or not -- process.

You evaluate not only individual behaviours but also organisations, material and immaterial resources, services and all that is <<human>>. Evaluation process can be implicit or explicit. If explicit - therefore communicable - it can use numerical values (scores) or verbal descriptions (judgements). During the last few years a great interest about new evaluation methods has been increasing: from school to companies, from market to financial analysis, from short to long range and strategic planning...

In fact the traditional methods used are criticised because too simplistic or disorganised to <<catch>> and evaluate complex, difficult or ambiguous situations. For all these cases evaluation systems based on Fuzzy Logic are being developed. These methods enrich information of contents and allow a less conflicting, but more rapid and complete synthesis of judgements.

2. Traditional evaluation scales

For surveys with traditional methods, two kinds of scales are usually used:

- *numerical scales:* they allow people to answer a question, expressing a score related to a pre-defined judgement.
  For example: to judge the height of Mister X respectively with the score 5, 6 or 7, if you think his height is: less than 150 cm, between 150 and 170 cm or more 170 cm. (Fig. 1 -- omitted)
- *numerical-verbal scales:* you ask to give a pre-defined judgement on an increasing (or decreasing) verbal terms scale. (Fig. 2 -- omitted)
  From a theoretical point of view there are no differences between the two kinds of scales: in fact each judgement, numerical or numerical-verbal is exactly defined. But both exactness of definition and preciseness of scales are only apparent. In fact if you translate the height of three people A, B, e C who are respectively a = 149 cm, b = 151 cm e c = 169 cm tall, the result is (figure 3 -- omitted):
- Mister C and Mister B are considered <<medium>> tall even if they differ of 18 cm;
- Mister A is judged irremediably <<short>>, but Mister B, who is taller that him of only 2 cm, is judged <<medium>> tall.

But there is something more: if evaluators use common sense as judgement measure, very probably they will neglect a few centimetres difference and consider Mister A and Mister B as members of the same category (they will be judged both short or medium) specially referring to Mister C.
Besides, evaluators' calibration, even if there is a pre-fixed scale, is subjective: an evaluator about 2 metres tall can consider short A, B and C. This is a case similar to that of a twenty years old person who is judged old from a child aged 5, but young from person aged 90.

Finally, calibration of judgements can be conditioned by the purpose of the evaluation:
Mister C, 169 cm tall, can be considered too short to play basket or too tall to become a professional house-jockey.

From what has been said above, some conclusions follow:

- the subjectivity and conditioning of evaluators have to be recognised and, possibly, explicated: evaluations are always subjective and evaluators have the responsibility to ensure evaluations not to be tendentious;
- usually used scales have some fundamental faults:
  - *discreet scales:* a term of these scales may include very different features (in the example Mister B and Mister C heights are considered equal);
  - *non-continuos scales:* two contiguous terms (short and medium) underline differences which are inferior to others which are neglected (Mister A and B's heights).
  - *pre-defined scales.* A self confident evaluator can express the certainty of his evaluations only substituting the pre-fixed scale with a more detailed one; this isn't usually allowed by questionnaires (vice versa a less self confident evaluator can't downsize the terms of scales).

It's simple to notice that, in everyday life, evaluations are done using different scales, customising them to circumstances and evaluators.

A clear example is the evaluation in Italian high schools. In fact professors, even if a 11 termed numerical scale (from 0 to 10) is available, generally they
3. use reduced scales: the most common starts from 3 to 4 and reaches 8,
4. divide scales using signs (+), (++), (-), (--), etc.
5. use asymmetric scales: scales have, generally, more positive scores than negative ones with thicker intervals around 5.

This corresponds to the evaluators' necessity to customise the evaluation scale, in order to give to the rater a more adherent and precise judgement, according to facts estimated as important and to the certainty of their estimate. So in schools, as in everyday life, scales with less terms are used when a fact is not very known, not clear or when there isn't an exact opinion about it. On the contrary scales become more precise, that is with more terms, each one with a reduced range, when the certainty of judgement allows to better clarify the evaluation,

3. Verbal Judgements

As it is known, in many cases, instead of the scales above, verbal approaches are used; that is evaluations are expressed in a natural language. In this way it is possible to formulate more complex, fading and even ambiguous judgements.

> <<The vagueness of the words, that is strictly linked to the process of human categorisation of the experiences, allows the grouping of many different, even contradictory facts>> (Russel, 1923).

Yet the verbal approach underlines notable limits especially when synthesising more judgements. For example:

- the efficiency of the synthetic process is limited by the need to decode many propositions, even complex ones;
- when there are difficulties in the process co-ordination, an organisation tends to decentralise it. So divided entities expressing conflicting judgements can be formed,
- final syntheses loose the richness of each person and groups and they aren't completely transparent and clear to be decoded from evaluators (or from the evaluated person);
- in the process of synthesis not the most expert, but the most authoritative (or aggressive) person can prevail, even if he/she is less prepared on the specific situation to evaluate.

3. <<Fuzzy>> evaluation scales

To cross the problems of discrete and discontinuous scales, fuzzy ones allow a superimposition of their terms. For example the scales above (figg. 1 and 2) are substituted by the one below in figure 4(omitted):
In this way it is possible that individuals between, for example, 145 and 160 cm can be considered short and medium from a part of evaluators. To better describe this possibility, each term of the scale (for example: medium) is characterised by a membership distribution, of people evaluated, to the same term. In our case the distribution is described in a more simple way, by a triangular function, as in figure 5 (omitted).

Referring to the figure we can say that M is surely short (M = 144 cm), but also N (N 146 cm) will be considered not really a member of the set <<medium>>. In fact if you express the possibilities of membership in percentage it will result that P belongs 100% to <<medium>> term, while N belongs to it only 6,7%. You can say that, being 100 the population of observers, only the 6 or 7% will consider N as medium, while the others will consider it short as M.

To overcome the limit of pre-defined scales, <<fuzzy>> questionnaires offer, for each judgement, several scales, varying in the number of terms (Term Set = TS). In annex 1 a questionnaire with 11 scales is drawn:

- from TS1, which, with only a term, makes clear the evaluator's inability to evaluate, or the weak will to express himself,
- to TS 11 which makes clear the maximum certainty in judging, the best knowledge of the object under evaluation and all the availability to get exposed.

It's clear that the a priori choice of a scale allows the evaluator to announce, in advance, his/her self-evaluation in relation to the judgement he/she wants to express. The further choice of the term on the selected scale can be detailed and motivated by propositions, useful to have a synthesis more easily shared with the other evaluators. It can also limit, at least, problems posed by pure verbal approaches.

The result of many evaluators' judgement is usually a <<cloud>> of judgements, represented by some verbal terms on some scales, supported by possible motivations.

The elaboration of judgements, in order to extract the <<average>>, was presented by Zollo G. and others in: << *Understanding and managing organisational paradox with fuzzy logic*)). IIIº SIGEF Congress 1996, Buenos Aires, 10^{th} - 13^{th} November.

It's based on the fact that each term of a <<fuzzy>> scale can be expressed with two numbers, obtained by crossing the <<gaussian>> curve of the same term (simplified with a isosceles triangle) with truth and untruth curves of the predicate, expressed by the range of the scale to which the term belongs.

Referring to the previous case and to figure 6 (omitted), the two values (0,6) are the indexes of the <<medium>> term, which is the second one in TS3 fuzzy scale, pointing out the predicate linked to height.
Fiat Research Centre, on these bases, developed a registered software (CRF/R.E.S.®) to quickly calculate <<fuzzy>> average and dispersion of judgements expressed from some evaluators.

In the following examples we demonstrate that it isn't always necessary to calculate the average of judgements.
In many cases, by the examination of graphic dispersions of scores and the analysis (if possible) of its motivations, it is possible to:

- agree on a synthetic shared position (in a much more easy way than with standard scales);
- decide measures on a <<common sense>> average.

4. Examples of <<fuzzy>> questionnaires applications

- The examples are aimed at:
- demonstrating versatility of application;
- suggesting work of research and tuning on questionnaires, that can and must be done to fully exploit their potential.

4.1 Market survey on grocery good's <<packaging>>

Annex 2(omitted) is summarising a survey on 112 customers, in which a company asked some judgements on the features (A, B, C,...) of one of its products *(K)* and about their expectations on the product category (to which *K* belongs) *XYZ*. For a clearer explication we will detail the results of question <<A>>, proposed for the product and its category.

> Question A1: Do you think the *packaging* of *XYZ* products (to which *K* belongs) is important to guarantee their features during the time?
>
> Question A2: Do you think that product *K* has a *packaging* that satisfies your expectations?

In Annex 2 comparing, for the item A (<<packaging>> features), <<fuzzy>> averages of consumers expectations to product performance, we have a difference which can be judged not very important. In fact the two judgements of synthesis are quite positive and with a relatively weak difference.

But if you compare judgement distributions in the following tables you can formulate many more considerations:
From annex 3 (omitted) you can, for example, say that:

- consumers are very sure in their assertions: their most used scale is TS 11 and only few people use scales inferior than TS5;
- scores, not so negative, are only 5 and neutral ones are 13, totally 16% of not positive judgements.

Instead, annex 4(omitted) indicates that:

- consumers are just a little more uncertain (or less available to be exposed) because the majority of them uses scales just a little inferior;
- however, scores are much more dispersed: in fact, negative and very negative scores are 33, with 23 neutral ones; only 50% of evaluators think product *K* packaging is adequate.

The analysis of justifications about scores leads the company of *K* to quickly improve its packaging.

4.2 Survey on validity of new evaluation methods

This questionnaire was given to 31 professors of ITC Marro in Moncalieri, in order to verify their aptitude to use new methods for evaluation in schools.

Professors could choose among 8 scales, from 2 (TS2) to 9 (TS9) terms, according to the certainty of their judgement (annex 5-- omitted). Professors were very sure: in fact only 3 of them used scales with few terms. The average result, calculated with fuzzy mathematics, is the 7th term of TS8 scale, that is <<very useful, quite indispensable>>. Even without the mathematical calculation, you can see that scores are concentrated in the right superior quadrant. Simply dividing the sheet in four quadrants you have the indications of figure 7 (omitted).

4.3 ITC Marro Core competencies evaluation

In annex 6(omitted) ITC Marro's core competencies evaluation is proposed. It was made by a statistical sample composed by professors (60) and students. In particular this annex shows students and professors evaluations averages: generally students are less sure in judging (scales with few terms) and more careful (judgements less positive).

In annex 7(omitted) and annex 8 (omitted) there are the scales of judgements for each competency (A, B, C, ecc..) with their average score and scores dispersions graphics (distance between each score and the average value) of professors and students. Groups of competencies were classified according to evaluation values and dispersions as points of strength or critical ones at different levels.

Comparing points of strength and critical ones expressed by professors and students, improvement and optimisation points were found out: they are all synthesised in annex 9 (omitted).

4.4 Survey on a company canteen

The judgements on a canteen in a company are shown in annex 10 (omitted). Evaluation scales available were only 3 (TS7 - TS5 - TS3). The not positive result was visible with the quadrants technique already described. <<Fuzzy>> evaluation confirms the negative evaluation (faded), pointing out the 3rd term <<Lower Average)) as average value of a 6 termed scale.

4.5 Individual competencies evaluation

An example of company people evaluation is shown in annex 11(omitted). The requested evaluation is based on a numerical -verbal scale, presenting defects pointed out above in the first paragraph. In fact 6 <<crisp>> terms are admitted. The first three of them concern the existence or not of knowledge. The last three terms define competencies as: knowledge + results.

In annex 12 (omitted) a <<fuzzy>> sheet is used in order to get a better description of the person evaluated and to allow evaluators to be freer in their expression. It shows 4 scales that can be projected on the standard numerical-verbal scale. These scales are divided in two halves: in the first one, terms indicate only the ownership of knowledge; in the second one it's requested to estimate the level of competency, i.e. knowledge + the capacity to apply it + the capacity to make people apply it in order to reach results in work situations. Therefore we can say that the difference in the evaluation process is due to results. If they lack or they are negative, evaluators can use only the first half of the scale and, for this half, a further difference is given by the level of participation. The evaluator chooses the scale and then one of its terms, that can be enriched indicating the most important work situations, in which he/she observed the person under evaluation in action: that is when he/she is applying knowledge or competency under exam.

Conventionally starting from a <<fuzzy>> score you arrive to a numerical-verbal scale, projecting on it the average value of the term chosen. Notice that the most uncertain scale (TS4) doesn't allow extreme scores.

5. Conclusions

What has been presented pertains to the field of <<fuzzy>> scales applications, first graphically drawn and then experimentally used. Naturally, both the tool and its representation can be improved. The positive fact is, then, given by the possibility to apply simple methods to obtain meaningful results for the users and to enrich the method more and more, as the evaluators' ability increases. In particular the last tool presented, for individual competencies evaluation, is an extremely simplified version of a much more complex approach, which integrates pre-defined competencies evaluation (or Top-Down) with the freely expressed ones by evaluators (Bottom-Up). This methodology, based on Fuzzy logic, was developed to answer evaluation needs of dynamic, knowledge intensive companies, in turbulent environments; that is for companies which will be successful in the near future.

References

Baas, Kwakernak, Rating and ranking of multiple aspect alternative using Fuzzy Sets, *Automatica,* vol. 13, p. 47-58.

Barley, (1983), Semiotic and the Study of Occupational and Organizational Cultures, *Administrative Science Quartely,* n 28, p. 393-413

Berger, Luckmann, (1966) *The social Construction of Reality: A Tratise in the Sociology of Knowledge,* New York, Duobleday.

Cannavacciuolo A,, Capaldo G., Michellone G.C., Zollo G., (1996), "Understanding and Managing Organizational Paradox with Fuzzy Logic", III SIGEF Congress, Buenos Aires, November 10-13.

Zadeh L.A., (1996), Fuzzy Logic = Computing with Words, *IEEE Trans. On Fuzzy Systems,* vol.4, N. 2, p. 103-111.

Zollo G., Cannavacciuolo A., Capaldo G., Ventre A., Volpe A., (1996), The organizational evaluation process: a fuzzy model, *Fuzzy economic Review* nº 1, Vol. I, May 1996.

About the Author
GianCarlo Michellone, President and CEO of CRF (FIAT Research Center)

Contact details:

Ing. GianCarlo Michellone
CRF
Strada Torino 50
I-10043 Orbassano (Torino), Italy