Generalized evaluations in sociological researches

Abdirimov Kudrat Raximberganovich
Department of software of information technologies
Tashkent university of information technologies(TUIT)
Tashkent, Uzbekistan
dasturchi@mail.ru

Turimov Dilmurod Mustapoyevich
Department of software engineering
Samarkand branch of TUIT
Samarkand, Uzbekistan
turimov86@gmail.com

Abstract—Considered the calculation of generalized evaluations in heterogeneous characteristic space. Methods of data mining are used for calculating generalized evaluations.

Keywords-*experts,generalized evaluations,attributes,gradation, classess*

*****

I. INTRODUCTION

The use of generalized evaluations in sociological researches has been significant in the field. It is possible to determine the opinion of people about particular events, actions or happenings, their intelligence quotient and how different social and demographic groups receive information via generalized indicators [1]. Up to now, generalized evaluations were carried out only by expert judgment (expert querying).

It is generally accepted, experts who have expertise in a particular field are able to check generalized evaluations and they are the following [2]:

- Points of experts;
- Organizing the level of exposure of analyzed features of investigated objects by experts;
- Expert estimations in even comparison matrix view.

The drawback of above mentioned expert assessment method is revealed in the following:

Determination of informative collection of factors conveyed in different measurement units has not been proved by theoretical-methodological relation.The reliance of experts on subjective ideas when evaluating generalized evaluation.

By the method suggested in the paper, the calculation of the generalized evaluations is done via the intellectual analysis method. On the basis of the calculated generalized evaluation, latent (hidden) knowledge is developed.

The objective of the expert is to analyze information on the basis of "object-property" table via this method. When calculating the generalized evaluation, the unrelieved information is found on the basis of intellectual analysis of artificial neural network technology on the table [3].

The method is used to calculate the generalized evaluation and to explain how they are identified. In the following database, there is a method to find hidden rules.

Experts have an opportunity to check their hypothesis of considered matters on the basis of the method.

Experts may give an account of acquired knowledge as a result of investigation based on linguistic rules, if necessary they may state it by a clear formula.

II. THE SOLUTION OF THE PROBLEM

The objective is to determine the function $y = f(x_1,\ldots,x_n)$ by the value of cause indicators.

The problem of identification in standard setting is consedered. The multiplicity of objects $E_0 = \{S_1,\ldots,S_m\}$, containing representatives of 2 uncrossing classes K_1, K_2 has been stated.The description of objects is fullfilled with the help of sets $X_n = (x_1,\ldots,x_n)$ fromdiverse features n , where ξ of those (a subset $X(I) = (x_1,\ldots,x_\xi)$) are measured in interval scales, $n-\xi$ (subset $X(N) = (x_{\xi+1},\ldots,x_n)$) – in nominal $X_n = X(I) \bigcup X(N)$.

It is possible to change any k (k>2) class problem into the form of 2 class problems.

The necessity to review the objectives of two class problems is based on the following:

- firstly, any generalized mark (indicator) has a relative feature. Any class objects are set contrary to other class objects. For instance, the opinion of the class members on choosing a spouse who have higher education degree and who does not have;

- Secondly, there are no analytical functions to restore connections in space of features of different types classes.

III. THE CALCULATION OF GENERALIZED EVALUATIONS [1]

The weight of value features is found by separating objects into K_1, K_2 classes.

The feature value of x_j which is set in increasing progression is separated into two intervals $[c_1,c_2],(c_2,c_3]$ and each of them is considered as nominal gradation feature. The border setting of c_2 is based on the hypothesis which states that each period includes value features of only one class.

We shall consider the number of symbol value of $x_j, j \in I$ as u_i^1, u_i^2 which belongs to class $K_i, i=1,2$ and in the interval $[c_1,c_2],(c_2,c_3]$.

p - is the determination of element's sequence number which is arranged in increasing progression of the feature values of x_j from $r_{j1}, r_{j2}, \ldots, r_{jm}$ and in the border interval of $c_1 = r_{j_1}, c_2 = r_{j_p}, c_3 = r_{j_m}$.

The following criterion creates an opportunity to calculate optimal values of borders between intervals $[c_1, c_2], (c_2, c_3]$ and to express their quantitative features in nominal measurement.

$$\left(\frac{\sum_{i=1}^{2} u_i^1(u_i^1-1)+u_i^2(u_i^2-1)}{\sum_{i=1}^{2}|K_i|(|K_i|-1)}\right)\left(\frac{\sum_{d=1}^{2}\sum_{i=1}^{2} u_i^d(|K_{3-d}|-u_{3-i}^d)}{2|K_1||K_2|}\right) \to \max \tag{1}$$

In the above given formula, the expression in the left bracket expresses the resemblance in the class and in the right bracket, the difference among classes.

$(i \in I)$ is the optimal value of i - attribute therein, according to the criterion w_i (1). c_1^i, c_2^i, c_3^i is the separating interval border which is equivalent to the value as well as corresponding number of values of K_1 and K_2 class in $[c_1^i, c_2^i]$ interval of u_{i1}^1, u_{i2}^1 criterion.

The determination of generalized evaluation of objects by quantitative attributes is as following. In order to calculate the value of the object $S = (x_1, \ldots, x_n)$, the following function is used:

$$R(S) = \sum_{i=1}^{n} w_i t_i (x_i - c_2^i)/(c_3^i - c_1^i),$$

The value of $T = (t_1, \ldots, t_n)$ set is found by the following condition:

$$\min_{S_p \in K_1} R(S_p) - \max_{S_p \in K_2} R(S_p) \to \max.$$

$t_i \in \{-1, 1\}$ therein.

It is necessary to identify the weights of nominal attributes and their gradation for the objects expressed by various types of attributes.

The weight of nominal attributes and determination of their gradation:

r - number of attribute gradation is determined via *p*. $r \in J$ therein. g_{dr}^t is the number of gradation t belonging to the class K_d of r – attribute. ($1 \le t \le p$)

l_{dr} is the number of gradation of r - attribute in K_d class.

The distinction between K_1 and K_2 classes is determined according to r - attribute as following:

$$\lambda_r = 1 - \frac{\sum_{t=1}^{p} g_{1r}^t g_{2r}^t}{|K_1||K_2|}. \tag{2}$$

The value of β_r of the equality level of K_1 and K_2 classes in r - attribute is found by the following formula:

$$D_{dr} = \begin{cases} (|K_d| - l_{dr} + 1)(|K_d| - l_{dr}), & p > 2 \\ |K_d|(|K_d| - 1), & p \le 2 \end{cases}$$

$$\beta_r = \frac{\sum_{t=1}^{p} g_{1r}^t (g_{1r}^t - 1) + g_{2r}^t (g_{2r}^t - 1)}{D_{1r} + D_{2r}}. \tag{3}$$

$v \in J$ weight of the nominal attribute is identified as stated below on the basis of the above-given (2) and (3) formulas:

$$v_r = \lambda_r \beta_r. \tag{4}$$

It is easy to check if the nominal attribute weight is between interval [0, 1].

Evidently, it is possible to express gradations of nominal attribute in the form of $\{1, \ldots, p\}$ set. When calculating the generalized evaluation for object $S = (x_1, \ldots, x_n)$, $x_i = j$ attribute fraction is determined as in the following($i \in J$, $j \in \{1, \ldots, p\}$)

$$\mu_i(j) = v_i \left(\frac{\alpha_{ij}^1}{|K_1|} - \frac{\alpha_{ij}^2}{|K_2|} \right),$$

$\alpha_{ij}^1, \alpha_{ij}^2$ is the number of j – gradation in i – attribute in K_1 and K_2 classes.

v_i is the weight of i – attribute calculated according to the formula (4).

The generalized evaluation of each $S_a \in E_0$, $S_a = (x_{a1}, \ldots, x_{an})$ object which consists of nominal attributes is found by the following formula:

$$R(S_a) = \sum_{i \in I} w_i t_i (x_{ai} - c_2^i)/(c_3^i - c_1^i) + \sum_{i \in J} \mu_i (x_{ai}) \tag{5}$$

Generalized evaluation (5) is closely associated with the value of the weight of quantitative (1) and nominal attributes (4).A conclusive logic is used to express the generalized evaluation of E_0 objects in interval [0, 1].

For instance, test value according to generalized evaluation of E_0 objects shows mixture level of objects of K_1, K_2 classes. The higher this value is, the lower its mixture level. Accurate (error-free) separation of classes (1) occurs when the value equals to one.

In this case (5), a discriminant analysis of information as a critical function may be suggested.

IV. COMPUTING EXPERIMENT AND CONCLUSIONS

248 subjects have been used in the research. 1-class includes literate people of Pakistan, 2-class, 124 literate people of Afghanistan. In the contest 33 nominal attributes (question) have been used. The informational level of each attribute (question) has been identified by formula (4) and arranged in digressive order in Table 1.

TABLE I. THE LEVEL OF INFORMATION TRANSFER OF ATTRIBUTES

№	Attribute	Weight	Difference among classes	Similarities within class
1	Has there been built a local radio station in your location?	0,481	0,645	0,746
2	Place of residence	0,362	1	0,362
3	If a local radio station has been built in your area, does it meet requirements set by you?	0,362	0,851	0,425
4	National news.	0,348	0,591	0,589
5	Local news.	0,347	0,568	0,612
6	International news.	0,34	0,584	0,582
7	Religious manuals.	0,325	0,56	0,58
8	What is your opinion about how different local radio from national and international ones?	0,315	0,922	0,342
9	What is your sex?	0,312	0,56	0,557
10	Cultural events.	0,288	0,5	0,576
11	Type of school.	0,284	0,5	0,569
12	Do you discuss with your parents?	0,274	0,43	0,638
13	Formal and unofficial education.	0,262	0,503	0,522
14	Entertaining programs.	0,261	0,512	0,509
15	Do you discuss with neighbors?	0,256	0,502	0,511
16	Do you discuses with friends?	0,254	0,414	0,615
17	What is your field of interest in obtaining information?	0,253	0,634	0,399
18	Why do you listen to the radio?	0,252	0,682	0,369
19	Region. (city or countryside)	0,249	0,491	0,507
20	What is your favorite radio program?	0,235	0,795	0,296
21	What is the source of information in your area?	0,231	0,31	0,745
22	What radio do you listen the most?	0,23	0,78	0,295
23	Do you discuss with your spouse?	0,217	0,309	0,702
24	Do you discuss with children?	0,216	0,317	0,681
25	Why do you specifically favour this program?	0,187	0,801	0,234
26	Why do you specifically favour this radio?	0,185	0,81	0,2274
27	Native language.	0,164	0,192	0,853
28	What programs do you like discussing?	0,163	0,849	0,1911
29	What is your age?	0,156	0,832	0,1874
30	What is your job?	0,138	0,865	0,1591
31	Do you discuss with other people?	0,125	0,145	0,862
32	I do not discuss with anybody.	0,098	0,107	0,914
33	And other programs.	0,079	0,086	0,916

The rate of questions, their difference among classes and likeness in the class is between interval [0, 1]. The rate of questions occurs from multiplication of differences among classes and likeness in the class. If the value of likeness in the class is 1, it means that the members of the class answered unanimously. If the value draws to 0, it is assumed that class members have different opinions on the attributes (questions) given to them. If the value equals to 1 among classes, both class members have answered to questions differently from each other. If the value is draws closer to 0, both class members provided similar opinions.

The difference among classes is 1 in the 2^{nd} question "Place of residence" in Table 1. It means each member of class has answered differently from members of class 2. That is, they reside in different countries. The similarity in the class (Pakistan 0.173; Afghanistan 0.189) means that respondents are from different regions.

TABLE II. THE GENERALIZED EVALUATION OF OBJECTS BELONGING TO K_1 CLASS.

№	Object (class)	R(S)	№	Object (class)	R(S)
1	70^{th} object ("state")	1,0	129	199- object ("non-state")	0,40
2	69- object ("state")	0,99	130	184- object ("non-state")	0,40
3	74- object ("state")	0,99	131	172- object ("non-state")	0,39
4	77- object ("state")	0,95	…	…	..
5	72- object ("state")	0,93	160	10- object ("state")	0,28
...	...	...	161	11- object ("state")	0,28
58	228- object ("non-state")	0,69	…	…	..
59	31- object ("state")	0,69	245	234- object ("non-state")	0,02
…	…	..	246	221- object ("non-state")	0,02
127	51- object ("state")	0,41	247	214- object ("non-state")	0,01
128	13- object ("state")	0,40	248	220- object ("non-state")	0,00

Definitely, it is possible to draw the following conclusions as a form of knowledge based on the obtained results in the case that the qualitative explanation of the generalized evaluation of the objects is in the authority of field experts:

The value of objects 70, 69, 74, 77, 72 belonging to "state" class of the selection almost equals to 1 and therefore they are explicit examples of the class. The same goes for objects 220, 214, 221, 234 of "non-state" class which are examples of it and their value belonging to "state" class is nearly 0.

51,13 object "state" class of the selection is considered as boundary elements. Their belonging to the class is revealed relatively slowly.

Belonging to their own class is revealed much slowly in the objects 10,11 of the "state" class in the selection.

228 object of "non-state" class has some belonging to "state" class.

ACKNOWLEDGMENT

A database of sociology laboratory of Uzbekistan National University is used as a calculation experiment.

REFERENCES

[1] Tolstova Y.N. "Analysis of sociological data. Descriptive statistics, the study of relations among nominal attributes". Scientific world, Moscow, 2000.

[2] Ayvazyan S.A., Bukhshtaber V.M., Yenyukov I.S., Meshalkin L.D. "Applied statistics". Finances and statistics, Moscow, 1989.

Article in a journal:

[3] Ignatev N.A. "Computation of generalized indicators (evaluations) and data mining". Automation and telemechanics. 2011. №5. Pp.183-190.