ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION

The method of assessing the quality of work of the Commission to verify exams

Гранков Михаил Васильевич, кандидат технических наук, доцент;
Щербинин Тимофей Александрович, аспирант;
Краснова Евгения Витальевна, кандидат филологических наук, доцент;
Донской государственный технический университет (г. Ростов-на-Дону)

The paper presents a method of evaluating checking quality of a Unified State Exam by an expert commission.
Key words: *fuzzy logic, cluster analysis, evaluation of the quality of work, USE.*

Introduction. Since 2009 in the Russian Federation, the exam in the form of the Unified State Exam (USE) has become the only form of general certificate of secondary education and at the same time university entrance exams.

At this stage, an ambiguous attitude to this form of the exam has developed in the society. In this connection the task for examinations on the subjects are being continually updated and changed.

For most subjects, all the tasks are divided into three groups, depending on their complexity. In this case, the answers to the task of "group A" (with a choice of possible answers) and "group B" (short answer) should be filled in special form filled in a strict accordance with the submitted letters to fill, since verification is done on a computer. Assignments of the "group C", requiring a detailed answer, should be written on a special form and checked by experts of the regional certification Commission. It should be noted that there are cases of incorrect scoring by the experts when checking "group C" tasks. In this regard, the final mark obtained by the examinees, depends not only on the correctness of the tasks done during the exam, but also on the competence of the inspection expert who can prevent a number of errors when checking the tasks. A school leaver may also appeal the results of the evaluation to a special commission.

Thus, an important goal is the problem of the expert committee formation, checking assignment of the "group C".

Error checking examination papers. The expert verification procedure is normalized by special instructions of "Federal Institute for Education Measurement" (FIEM), which contain detailed algorithms of each task evaluation for each task variants. To implement the evaluation each expert is provided with the scanned real work copies of the school leavers.

To reduce the effect of errors in the expert evaluation results for each specific "group C" task verification procedure is provided by two experts of the subject or another regional commissions. If the assessment of these experts are substantial by different, the major of them is taken into account. For each subject there is its own criterion for estimates slight difference. If the experts assessments are not close enough according to a specific criterion for the subject, then this task will automatically be sent to the third expert, whose decision will be final [1].

Assessment methodology. The developed method evaluates the quality of the experts work according to the number of the checked tasks. The analysis of the amounts on the revision expert, work of these problems is done.

Three groups were used for evaluating the quality of the experts, "low (low check quality)", "average" (average check quality) and "high" (high check quality).

Because of the multiple experiments at various examinations, the commission concluded that the key indicators of the work performance evaluation are the frequency adjustments for 3-point inspection and specific evaluation, which plays an important role if the indexes adjustments frequency of the experts are close to each other.

Frequency adjustment points at 3 checks d_i^{e3} for the i-th expert, will be calculated by the formula (1):

$$d_i^{e3} = \frac{t_3}{m_i}, \quad (1)$$

where t_3 — the number of errors made in 3-rd check;

m_i — the number of tasks, checked by the i-th expert.

Alternative points ($xalt3_i^k$) for the i-th expert to specify k, which was the third test are equal to the score, which was exhibited by the 3rd expert:

$$xalt3_i^k = x_3^k, \quad (2)$$

where x_3^k — scores of the 3-rd expert for the k-th job (can be omitted).

The absolute value calculation of the estimations deviation ($xabs3_i^k$) for the i-th expert in the k-th task with the third check, we can find with the formula (3):

$$xabs3_i^k = \left|xalt3_i^k - x_i^k\right|, \quad (3)$$

where x_i^k — scores for the i-th expert for the k-th job;

$xalt3_i^k$ — alternative score for the k-th job by formula (2).

Specific variations in the absolute value of 3 checks ($xabst3_i^k$) we find with the formula (4):

$$xabst3_i^k = \frac{\sum_{k=1}^{m_i^3} xabs3_i^k}{m_i}, \quad (4)$$

where $xabs3_i^k$ — value deviation of the estimation by the i-th expert's job with number k by the formula (4);

m_i^3 — the number of tasks with a 3rd party checks of the i-th expert;

m_i — the number of tasks, checked by i-th expert.

Clustering method. As a method of clustering the estimated experts, which is combining them into groups according to the obtained results, the Fuzzy C-Means (FCM) algorithm was used [2]. This algorithm is most often used for fuzzy clustering data. It forms a group, prototypes of which are represented by points in the data space. To execute the algorithm on a set of data it is

ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION

necessary to choose the number of X groups, the fuzziness degree of n, the parameter ε in the stopping criteria of the algorithm, as well as randomly initiate a matrix of objects belonging to a particular group

$U^{(0)} \in Z$,

where i — the number of groups, and the prototype vector $V^{(0)}$.

As a condition to the completion of the algorithm (stopping criterion) quite small change in values of the elements of the matrix U is considered, i.e. $\|U^{(i+1)} - U^{(i)}\| < \varepsilon$, where ε — given constant [2].

The initial cluster centers are distributed evenly on the centers of the segments from the minimum value of the normalized parameter to the maximum. The centers of the clusters are selected to minimize the distance from the center to a group of objects, and to maximize the distance between the centers of the different clusters. Total number of clusters is assumed to be equal to three and is equal to the number of groups of experts estimating (low quality checks, the average quality of testing and high quality inspection).

The initial value of the center of the i-th cluster$centr_i$ () is given by:

$centr_i = \min(npar) + i * \frac{\max(npar) - \min(npar)}{2.0 * count}$,

where $\min(npar)$ — the minimum value of the normalized parameter of the sample;

$\max(npar)$ — maximum value of the normalized parameter of the sample; count — total number of clusters.

As a measure of distance used was weighted Euclidean metric was used according to which the distance ($dist$) from a point to a cluster is found by the formula:

$dist = \sqrt{k_1(x_c - x)^2 + k_2(y_c - y)^2}$,

where x_c, y_c — value coordinates of the center of the cluster,

x, y — the coordinates of the current value of the expert,

k_1, k_2 — weighting values of the estimated parameters [2].

Analysis of the obtained results. Let us consider the analysis of the expert committee by the example of mathematics.

It is obvious that in any work with the implementation of transactions tiredness and the likelihood of errors grows, so that all parameters of evaluation will be relate to the number of tested applications.

Since the adjustments are considered to be more important than the specific points of deviation, it was decided that their optimal ratio is 1 to 3.

Consider the example of the distribution of the group of experts according to the quality of evoluation. Abbreviations used in the tables are described in Table 1.

Table 1. **Abbreviations used**

N	expert number
PAR	frequency value adjustments in 3 checks (the total number of tasks)
PAR2	specific value of the absolute deviation score
PAR_N	normalized value of PAR
PAR2_N	normalized value of PAR2
G (%)	belonging percent to the cluster of high cluster validation
M (%)	belonging percent to the cluster of medium quality checking
B (%)	belonging percent to the cluster of bad quality checking
REIT	expert rating
BEL	resulting affiliation (calculated according to the more belonging percent): G (Good) good checking quality, M (Medium) the average checking quality, B (Bad) bad checking quality.

Table 2. **Evaluation of experts using a weighted Euclidean metric**

N	PAR	PAR2	PAR_N	PAR2_N	G	M	B	REIT	BEL
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0,53%	0,005333	0	0	97,53%	2,37%	0,10%	0,991	G
2	1,17%	0,011655	0,049	0,039	98,99%	0,98%	0,03%	0,944	G
3	1,29%	0,01476	0,059	0,058	99,24%	0,74%	0,02%	0,933	G
4	1,38%	0,013769	0,066	0,052	99,34%	0,64%	0,02%	0,928	G
5	1,72%	0,019084	0,093	0,085	99,75%	0,25%	0,01%	0,9	G
6	1,69%	0,037288	0,091	0,197	99,76%	0,23%	0,01%	0,89	G
7	2,13%	0,021277	0,125	0,098	99,95%	0,05%	0,00%	0,871	G
8	2,33%	0,027907	0,14	0,139	99,99%	0,01%	0,00%	0,853	G
...									
43	5,92%	0,073171	0,421	0,417	0,00%	100,00%	0,00%	0,574	M
44	5,96%	0,076596	0,424	0,438	0,00%	100,00%	0,00%	0,57	M
45	6,12%	0,082353	0,436	0,474	0,01%	99,99%	0,00%	0,555	M
46	6,38%	0,092199	0,457	0,534	0,11%	99,87%	0,02%	0,53	M
47	6,77%	0,075697	0,487	0,433	0,27%	99,65%	0,08%	0,515	M
48	6,69%	0,085502	0,481	0,493	0,22%	99,71%	0,07%	0,513	M
49	6,64%	0,097786	0,477	0,569	0,32%	99,56%	0,12%	0,508	M

ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION

1	2	3	4	5	6	7	8	9	10
50	6,85%	0,100806	0,494	0,587	0,56%	99,14%	0,29%	0,492	M
51	10,73%	0,138418	0,797	0,819	0,08%	0,62%	99,29%	0,199	B
52	10,95%	0,167883	0,814	1	0,03%	0,16%	99,81%	0,173	B
53	13,33%	0,16381	1	0,975	0,01%	0,06%	99,92%	0,009	B
54	13,33%	0,166667	1	0,993	0,02%	0,06%	99,92%	0,003	B

Figure 1 shows the distribution graphs of experts groups. Experts with bad checking quality are marked by squares, with an average quality- triangles, high - diamonds. Black dots on the graphs denote the center of the cluster. Sorting of experts goes on with the REIT parameter.

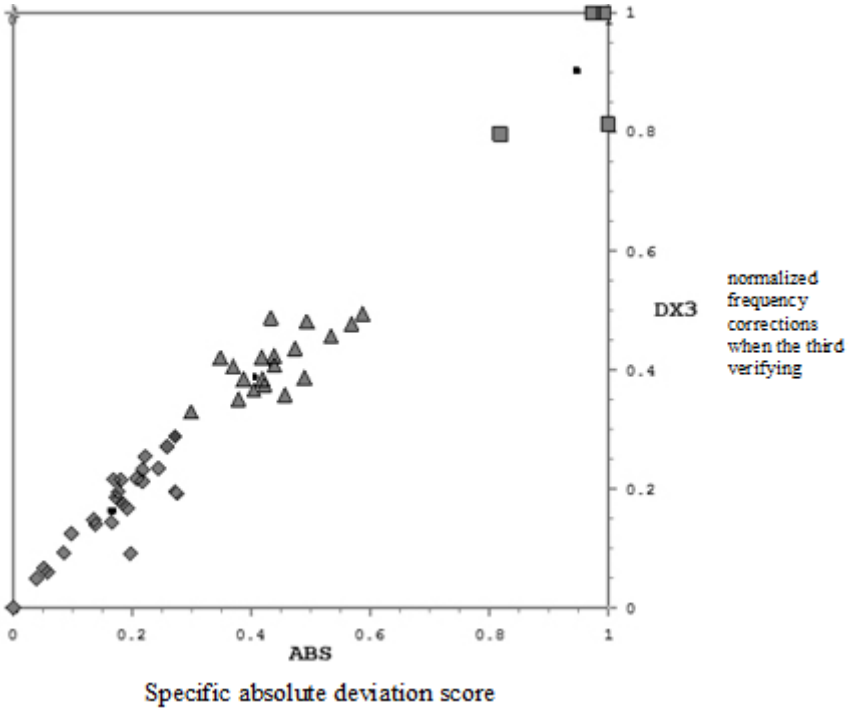

Figure 1. A graph of the expert groups

Conclusions. Our method allows to evaluate the quality of expert subject committees. Based on the obtained distribution by groups of experts, the work of regional examination commission can be improved. It is proposed if it is possible to form a committee for the next year in accordance with this distribution, refusing experts who are in a group of bad checking quality. Experts with average checking quality should be given recommendations about their work that will improve their testing indexes for the next year. Accredited tasks inspection is carried out by experts from the group with the highest quality, they are encouraged to be invited for a commission check. Experts from the group with good quality pro-verification assess the job well and they are encouraged to be invited to the committee for the following year.

References:

1. Гранков М.В., Щербинин Т.А. Оценка параметров работы экспертов предметной комиссии по проверке экзамена в форме ЕГЭ. Системный анализ, управление и обработка информации: Труды 5-го Международного семинара (п. Дивноморское, 2 – 6 октября 2014 г.) / Под общ. ред. Р.А. Нейдорфа. – Ростов-н/Д: ДГТУ, 2014. – С. 282-285.
2. Рутковский Л. Методы и технологии искусственного интеллекта. – Горячая Линия – Телеком, 2010. – 520 с.